

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:36:51 ; Search time 28.31 seconds
(without alignments)
9.066 Million cell updates/sec

Title: US-09-536-087-10
Perfect score: 53
Sequence: 1 WSPWAEW 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	53	100.0	1170	1	TSP2_BOVIN	Q95116 bos taurus
2	53	100.0	1172	1	TSP2_HUMAN	P35442 homo sapien
3	53	100.0	1172	1	TSP2_MOUSE	Q03350 mus musculus
4	50	94.3	1170	1	TSP1_BOVIN	Q28178 bos taurus
5	50	94.3	1170	1	TSP1_HUMAN	P07996 homo sapien
6	50	94.3	1170	1	TSP1_MOUSE	P35441 mus musculus
7	47	88.7	1173	1	TSP1_XENLA	P35448 xenopus lae
8	47	88.7	1178	1	TSP2_CHICK	P35440 gallus gall
9	45	84.9	470	1	PROP_CAVPO	Q64181 cavia porce
10	43	81.1	152	1	VG57_BPMD2	O64248 mycobacteri
11	43	81.1	152	1	VG57_BPML5	Q05270 mycobacteri
12	42	79.2	867	1	SSPO_BOVIN	P98167 bos taurus
13	41	77.4	421	1	TRAB_AGRU	Q44351 agrobacteri
14	41	77.4	574	1	CO9_ONCMY	P06682 oncorhynch
15	41	77.4	1074	1	SMSA_HUMAN	Q13591 homo sapien
16	41	77.4	1077	1	SMSA_MOUSE	Q62217 mus musculus
17	40	75.5	245	1	ATSA_BOVIN	Q9t193 bos taurus
18	40	75.5	387	1	TRAB_RHISN	P53416 rhizobium s
19	40	75.5	890	1	AT38_HUMAN	Q9up79 homo sapien
20	40	75.5	990	1	ENV_OMYVS	P16899 ovine lenti
21	40	75.5	1093	1	SMSB_MOUSE	Q60519 mus musculus
22	39	73.6	204	1	Y116_CAEEL	Q10919 caenorhabdi
23	39	73.6	320	1	STFQ_ECOLI	P77315 escherichla
24	39	73.6	557	1	CO9_RABIT	P48747 oryctolagus
25	39	73.6	559	1	CO9_HUMAN	P02748 homo sapien
26	39	73.6	803	1	FSPQ_XENLA	P35447 xenopus lae
27	39	73.6	807	1	FSPQ_RAT	P35446 rattus norv
28	39	73.6	837	1	ATSA_HUMAN	Q75173 homo sapien
29	39	73.6	909	1	Y4GI_RHISN	P55465 rhizobium s
30	39	73.6	967	1	ATSI_RAT	Q9wuq1 rattus norv
31	39	73.6	968	1	ATSI_MOUSE	P97857 mus musculus
32	38	71.7	140	1	LYSS_DROME	P37160 drosophila
33	38	71.7	336	1	G3P2_SYNY3	P80505 synechocyst

34	38	71.7	352	1	FMLR_RABIT	Q05394 oryctolagus
35	38	71.7	492	1	ADRO_BOVIN	P08165 bos taurus
36	38	71.7	586	1	CO9_FUGRU	P79755 fugu rubrip
37	38	71.7	860	1	AT56_HUMAN	Q9ukp5 homo sapien
38	38	71.7	879	1	YDBH_ECOLI	P52645 escherichla
39	38	71.7	905	1	AT58_MOUSE	P57110 mus musculus
40	37	69.8	140	1	CO8B_RAT	P55314 rattus norv
41	37	69.8	310	1	ARAC_ERWCH	P07642 erwinia chr
42	37	69.8	316	1	HEM2_DEIRA	Q9rv98 deinococcus
43	37	69.8	469	1	PROP_HUMAN	P27918 homo sapien
44	37	69.8	491	1	CCL_YEAST	P40986 saccharomyc
45	37	69.8	510	1	PUSH_ARATH	O22928 arabidopsis

ALIGNMENTS

RESULT 1
TSP2_BOVIN
ID TSP2_BOVIN STANDARD; PRT; 1170 AA.
AC Q95116; Q28180;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 2 PRECURSOR (CORTICOTROPIN-INDUCED SECRETED PROTEIN)
DE (CISP).
GN THBS2 OR TSP2 OR TSP-2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Danik M., Chinn A., Lafeuillade M., Keramidas M., Aquesse-Germon S.,
RA Penhoat A., Chen H., Mosher D., Chambaz E.M., Feige J.J.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 1-522 FROM N.A.
RX MEDLINE=96331130; PubMed=8698834;
RA Lafeuillade B., Pellerin S., Keramidas M., Danik M., Chambaz E.M.,
RA Feige J.J.;
RT "Opposite regulation of thrombospondin-1 and corticotropin-induced
secreted protein/thrombospondin-2 expression by adrenocorticotrophic
hormone in adrenocortical cells.";
RL J. Cell. Physiol. 167:164-172(1996).
[3]
RP SEQUENCE OF 318-831 FROM N.A.
RC TISSUE=Aortic endothelium;
RA Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
RT "Cloning and sequencing of bovine thrombospondin stimulatory effect of
TGF-beta.";
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
LAMININ AND TYPE V COLLAGEN.
CC -I- SUBUNIT: HOMOPRIMER; DISULFIDE-LINKED.
CC -I- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -I- SIMILARITY: CONTAINS 1 VFNC DOMAIN.
CC -I- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -I- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -I- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; X96540; CAA65385.1; -;
DR EMBL; X87620; CAA60952.1; -;

```

DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; VWFC.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 9.
DR Pfam; PF00093; vwc; 1.
DR Pfam; PF00181; EGF; 3.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50092; TSP1; 3.
DR PROSITE; PS01208; VWFC; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1170
FT DOMAIN 19 232
FT DOMAIN 318 375
FT DOMAIN 379 430
FT DOMAIN 435 491
FT DOMAIN 492 546
FT DOMAIN 547 587
FT DOMAIN 588 640
FT DOMAIN 646 695
FT DOMAIN 723 758
FT DOMAIN 759 781
FT DOMAIN 782 817
FT DOMAIN 818 840
FT DOMAIN 841 878
FT DOMAIN 879 914
FT DOMAIN 915 950
FT DOMAIN 951 1170
FT SITE 926 928
FT DISULFID 266 266
FT DISULFID 270 270
FT DISULFID 551 562
FT DISULFID 556 572
FT DISULFID 575 586
FT DISULFID 592 608
FT DISULFID 599 617
FT DISULFID 620 644
FT DISULFID 650 663
FT DISULFID 657 676
FT DISULFID 678 689
FT CARBOHYD 151 151
FT CARBOHYD 316 316
FT CARBOHYD 330 330
FT CARBOHYD 455 455
FT CARBOHYD 582 582
FT CARBOHYD 708 708
FT CARBOHYD 936 936
FT CARBOHYD 1067 1067
FT CONFLICT 535 535
FT CONFLICT 748 748
SQ SEQUENCE 1170 AA; 129862 MW; 9CF1FBF55B89A051 CRC64;

Query Match 100.0%; Score 53; DB 1; Length 1170;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSPNAEW 7
Db 382 WSPNAEW 388
|||||||

```

```

RESULT 2
TSP2_HUMAN
ID TSP2_HUMAN STANDARD; PRT; 1172 AA.
AC P35442;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 2 PRECURSOR.
GN THBS2 OR TSP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94010892; PubMed=8406456;
RA Labell T.L., Byers P.H.;
RT "Sequence and characterization of the complete human thrombospondin 2
cDNA: potential regulatory role for the 3' untranslated region.";
RL Genomics 17:225-229(1993).
RN [2]
RP SEQUENCE OF 560-1172 FROM N.A.
RC TISSUE=Fibroblast;
RX MEDLINE=92217961; PubMed=1559694;
RA Labell T.L., McGookey Milewicz D.J., Distche C.M., Byers P.H.;
RT "Thrombospondin II: partial cDNA sequence, chromosome location, and
expression of a second member of the thrombospondin gene family in
humans.";
RL Genomics 12:421-429(1992).
CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
LAMININ AND TYPE V COLLAGEN.
CC -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L12350; AAA03703.1; -.
DR EMBL; M81339; -. NOT_ANNOTATED_CDS.
DR PIR; A42173; A42173.
DR HSP; P00740; 11XA.
DR MIM; 188061; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; VWFC.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 9.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50092; TSP1; 3.
DR PROSITE; PS01208; VWFC; 1.

```

KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1172
 FT DOMAIN 19 232
 FT DOMAIN 318 375
 FT DOMAIN 381 432
 FT DOMAIN 437 493
 FT DOMAIN 494 548
 FT DOMAIN 549 589
 FT DOMAIN 590 647
 FT DOMAIN 648 692
 FT DOMAIN 725 760
 FT DOMAIN 761 783
 FT DOMAIN 820 842
 FT DOMAIN 843 880
 FT DOMAIN 881 916
 FT DOMAIN 917 952
 FT DOMAIN 953 1172
 FT SITE 928 930
 FT DISULFID 266 266
 FT DISULFID 270 270
 FT DISULFID 553 564
 FT DISULFID 558 574
 FT DISULFID 577 588
 FT DISULFID 594 610
 FT DISULFID 602 619
 FT DISULFID 622 646
 FT DISULFID 652 665
 FT DISULFID 659 678
 FT DISULFID 680 691
 FT CARBOHYD 151 151
 FT CARBOHYD 316 316
 FT CARBOHYD 330 330
 FT CARBOHYD 457 457
 FT CARBOHYD 584 584
 FT CARBOHYD 710 710
 FT CARBOHYD 1069 1069
 SQ SEQUENCE 1172 AA; 129955 MW; 2AC7BB230E4C6F5 CRC64;

 Query Match 100.0%; Score 53; DB 1; Length 1172;
 Best Local Similarity 100.0%; Pred. No. 0.79;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 WSPWAEW 7
 DB 384 WSPWAEW 390

 RESULT 3
 TSP2_MOUSE STANDARD; PRT; 1172 AA.
 AC Q03350;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE THROMBOSPONDIN 2 PRECURSOR.
 GN THBS2 OR TSP2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92147683; PubMed=1371115;
 RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
 RA Dixit V.M.;
 RT "Characterization of mouse thrombospondin 2 sequence and expression
 during cell growth and development.";
 RL J. Biol. Chem. 267:3274-3281(1992).
 RN [2]

RP SEQUENCE OF 1-873 FROM N.A.
 RX MEDLINE=91302287; PubMed=1712771;
 RA Bornstein P., O'Rourke K., Wikstrom K., Wolf F.W., Katz R., Li P.,
 RA Dixit V.M.;
 RT "A second, expressed thrombospondin gene (Thbs2) exists in the mouse
 genome.";
 RL J. Biol. Chem. 266:12821-12824(1991).
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 LAMININ AND TYPE V COLLAGEN.
 CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VWFc DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L07803; AAA53064.1; -;
 CC EMBL; M64866; AAA40432.1; -;
 CC PIR; A42587; A42587;
 CC PIR; A39851; A39851;
 CC HSP; P00740; IIXA.
 CC MGI; 98738; Thbs2.
 CC InterPro; IPR000561; EGF-like.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR001791; Laminin_G.
 CC InterPro; IPR000884; TSP1.
 CC InterPro; IPR003129; TSPN.
 CC InterPro; IPR001007; VWFc.
 CC InterPro; IPR003367; tsp_3.
 CC Pfam; PF00008; EGF; 2.
 CC Pfam; PF02210; TSPN; 1.
 CC Pfam; PF00090; tsp_1; 3.
 CC Pfam; PF02412; tsp_3; 9.
 CC Pfam; PF00093; vwc; 1.
 CC SMART; SM00181; EGF; 3.
 CC SMART; SM00209; TSP1; 3.
 CC SMART; SM00210; TSPN; 1.
 CC SMART; SM00214; VWC; 1.
 CC PROSITE; PS00022; EGF_1; FALSE_NEG.
 CC PROSITE; PS01186; EGF_2; 1.
 CC PROSITE; PS00092; TSP1; 3.
 CC PROSITE; PS01208; VWFc; 1.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1172
 FT DOMAIN 19 232
 FT DOMAIN 318 375
 FT DOMAIN 381 432
 FT DOMAIN 437 493
 FT DOMAIN 494 548
 FT DOMAIN 549 589
 FT DOMAIN 590 647
 FT DOMAIN 648 692
 FT DOMAIN 725 760
 FT DOMAIN 761 783
 FT DOMAIN 820 842
 FT DOMAIN 843 880
 FT DOMAIN 881 916
 FT DOMAIN 917 952
 FT DOMAIN 953 1172
 FT SITE 928 930
 FT DISULFID 266 266
 FT DISULFID 270 270
 FT DISULFID 553 564
 FT DISULFID 558 574
 FT DISULFID 577 588
 FT DISULFID 594 610
 FT DISULFID 602 619
 FT DISULFID 622 646
 FT DISULFID 652 665
 FT DISULFID 659 678
 FT DISULFID 680 691
 FT CARBOHYD 151 151
 FT CARBOHYD 316 316
 FT CARBOHYD 330 330
 FT CARBOHYD 457 457
 FT CARBOHYD 584 584
 FT CARBOHYD 710 710
 FT CARBOHYD 1069 1069
 SQ SEQUENCE 1172 AA; 129955 MW; 2AC7BB230E4C6F5 CRC64;

 Query Match 100.0%; Score 53; DB 1; Length 1172;
 Best Local Similarity 100.0%; Pred. No. 0.79;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 WSPWAEW 7
 DB 384 WSPWAEW 390

 RESULT 3
 TSP2_MOUSE STANDARD; PRT; 1172 AA.
 AC Q03350;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE THROMBOSPONDIN 2 PRECURSOR.
 GN THBS2 OR TSP2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92147683; PubMed=1371115;
 RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
 RA Dixit V.M.;
 RT "Characterization of mouse thrombospondin 2 sequence and expression
 during cell growth and development.";
 RL J. Biol. Chem. 267:3274-3281(1992).
 RN [2]

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7

Db 382 WSPWSEW 388

RESULT 5

```

TSPL_HUMAN
ID TSPL_HUMAN STANDARD; PRT; 1170 AA.
AC P07956;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 1 PRECURSOR.
GN THBS1 OR TSPL OR TSP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Endothelial cells;
RX MEDLINE=87057617; PubMed=2430973;
RA Lawler J., Hynes R.O.;
RT "The structure of human thrombospondin, an adhesive glycoprotein with
RT multiple calcium-binding sites and homologies with several different
RT proteins."
RL J. Cell Biol. 103:1635-1648(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89139590; PubMed=2918029;
RA Hennessy S.W., Frazier B.A., Kim D.D., Deckwerth T.L.,
RA Baumgartel D.M., Rotwein P., Frazier W.A.;
RT "Complete thrombospondin mRNA sequence includes potential regulatory
RT sites in the 3' untranslated region."
RL J. Cell Biol. 108:729-736(1989).
RN [3]
RP SEQUENCE OF 1-397 FROM N.A.
RX MEDLINE=87157592; PubMed=3030396;
RA Kobayashi S., Eden-McCuchan F., Framson P., Bornstein P.;
RT "Partial amino acid sequence of human thrombospondin as determined by
RT analysis of cDNA clones: homology to malarial circumsporozoite
RT proteins."
RL Biochemistry 25:8418-8425(1986).
RN [4]
RP SEQUENCE OF 1-374 FROM N.A.
RX MEDLINE=86287276; PubMed=3461443;
RA Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
RT "Characterization of a cDNA encoding the heparin and collagen binding
RT domains of human thrombospondin."
RL Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).
RN [5]
RP SEQUENCE OF 1-166 FROM N.A.
RX MEDLINE=89291870; PubMed=2544587;
RA Laherty C.D., Gierman T.M., Dixit V.M.;
RT "Characterization of the promoter region of the human thrombospondin
RT gene. DNA sequences within the first intron increase transcription."
RL J. Biol. Chem. 264:11222-11227(1989).
RN [6]
RP SEQUENCE OF 1028-1170 FROM N.A.
RA La Fleur M., Jobin C., Gauthier J., Kreis C.G.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-
CC V/BETA-3 AND ALPHA-ITB/BETA-3.
CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 WFEC DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.

```

```

-----
CC TH SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M25631; AAA36741.1; -
DR EMBL; X04665; CAA28370.1; -
DR EMBL; X14787; CAA32889.1; -
DR EMBL; J04835; AAA61178.1; -
DR EMBL; M99425; AAB59366.1; -
DR PIR; A05172; A05172.
DR PIR; A25812; A25812.
DR PIR; A26155; A26155.
DR PIR; A30140; A30140.
DR PIR; A34274; A34274.
DR HSP; P35555; IEMO.
DR MIM; 188060.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; WFEC.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 8.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50092; TSP1; 3.
DR PROSITE; PS01208; WFEC; 1.
DR Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1170
FT DOMAIN 19 232
FT DOMAIN 316 373
FT DOMAIN 379 430
FT DOMAIN 435 491
FT DOMAIN 492 548
FT DOMAIN 549 587
FT DOMAIN 588 645
FT DOMAIN 646 690
FT DOMAIN 723 758
FT DOMAIN 759 781
FT DOMAIN 782 817
FT DOMAIN 818 840
FT DOMAIN 841 878
FT DOMAIN 879 914
FT DOMAIN 915 950
FT DOMAIN 951 1170
FT SITE 926 928
FT DISULFID 270 270
FT DISULFID 274 274
FT DISULFID 551 562
FT DISULFID 556 572
FT DISULFID 572 586
FT DISULFID 592 608
FT DISULFID 599 617
FT DISULFID 620 644
FT DISULFID 650 663
FT DISULFID 657 676
FT DISULFID 678 689
-----
THROMBOSPONDIN 1
HEPARIN-BINDING (POTENTIAL).
WFEC.
TSP TYPE-1 1.
TSP TYPE-1 2.
TSP TYPE-1 3.
EGF-LIKE 1.
EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
EGF-LIKE 3.
TSP TYPE-3 1.
TSP TYPE-3 2.
TSP TYPE-3 3.
TSP TYPE-3 4.
TSP TYPE-3 5.
TSP TYPE-3 6.
TSP TYPE-3 7.
C-TERMINAL.
CELL ATTACHMENT SITE (POTENTIAL).
INTERCHAIN (PROBABLE).
INTERCHAIN (PROBABLE).
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.

```

FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 84 84 T -> A (IN REF. 2, 3 AND 4).
 FT CONFLICT 523 523 T -> A (IN REF. 2).
 SQ SEQUENCE 1170 AA; 129412 MW; 69B3EDF5AE3A395E CRC64;

Query Match 94.3%; Score 50; DB 1; Length 1170;
 Best Local Similarity 85.7%; Pred.No. 2.2;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
 |||||
 Db 382 WSPWSEW 388

RESULT 6
 TSPL_MOUSE
 ID TSPL_MOUSE STANDARD; PRT: 1170 AA.
 AC P35441;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE THROMBOSPONDIN 1 PRECURSOR.
 GN THBS1 OR TSPI.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92128941; PubMed=1774063;
 RA Lawler J., Duquette M., Ferro P., Copeland N.G., Gilbert D.J.,
 Jenkins N.A.;
 RT "Characterization of the murine thrombospondin gene.";
 RL Genomics 11:587-600(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92147683; PubMed=1371115;
 RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
 Dixit V.M.;
 RT "Characterization of mouse thrombospondin 2 sequence and expression
 during cell growth and development.";
 RL J. Biol. Chem. 267:3274-3281(1992).
 RN [3]
 RP SEQUENCE OF 1-490 FROM N.A.
 RX MEDLINE=90375546; PubMed=2398070;
 RA Bornstein P., Alfi D., Devayalu S., Framson P., Li P.;
 RT "Characterization of the mouse thrombospondin gene and evaluation of
 the role of the first intron in human gene expression.";
 RL J. Biol. Chem. 265:16691-16698(1990).
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-
 V/BETA-3 AND ALPHA-IIB/BETA-3.
 CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-1 DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M62470; AAA50611.1; -.

DR EMBL; M62450; AAA50611.1; JOINED.
 DR EMBL; M62451; AAA50611.1; JOINED.
 DR EMBL; M62452; AAA50611.1; JOINED.
 DR EMBL; M62453; AAA50611.1; JOINED.
 DR EMBL; M62454; AAA50611.1; JOINED.
 DR EMBL; M62455; AAA50611.1; JOINED.
 DR EMBL; M62456; AAA50611.1; JOINED.
 DR EMBL; M62457; AAA50611.1; JOINED.
 DR EMBL; M62458; AAA50611.1; JOINED.
 DR EMBL; M62459; AAA50611.1; JOINED.
 DR EMBL; M62460; AAA50611.1; JOINED.
 DR EMBL; M62461; AAA50611.1; JOINED.
 DR EMBL; M62462; AAA50611.1; JOINED.
 DR EMBL; M62463; AAA50611.1; JOINED.
 DR EMBL; M62464; AAA50611.1; JOINED.
 DR EMBL; M62465; AAA50611.1; JOINED.
 DR EMBL; M62466; AAA50611.1; JOINED.
 DR EMBL; M62467; AAA50611.1; JOINED.
 DR EMBL; M62468; AAA50611.1; JOINED.
 DR EMBL; M62469; AAA50611.1; JOINED.
 DR EMBL; M87276; AAA53063.1; -.
 DR EMBL; J05606; AAA40431.1; -.
 DR EMBL; J05605; AAA40431.1; JOINED.
 DR PIR; A40558; A40558.
 DR PIR; B42587; B42587.
 DR PIR; A37905; A37905.
 DR HSSP; P35555; IEMO.
 DR MGD; MGI:98737; Thbs1.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR000884; TSPI.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR001007; VWFC.
 DR InterPro; IPR003367; tsp_3.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF00090; tsp_1; 3.
 DR Pfam; PF02412; tsp_3; 8.
 DR Pfam; PF00093; vwc; 1.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00209; TSPI; 3.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50092; TSPI; 3.
 DR PROSITE; PS01208; VWFC; 1.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1170
 FT DOMAIN 19 232
 FT DOMAIN 316 373
 FT DOMAIN 379 430
 FT DOMAIN 435 491
 FT DOMAIN 492 548
 FT DOMAIN 549 587
 FT DOMAIN 588 645
 FT DOMAIN 646 690
 FT DOMAIN 723 758
 FT DOMAIN 759 781
 FT DOMAIN 782 817
 FT DOMAIN 818 840
 FT DOMAIN 841 878
 FT DOMAIN 879 914
 FT DOMAIN 915 950
 FT DOMAIN 951 1170
 FT SITE 926 928
 FT DISULFID 270 270
 FT DISULFID 274 274
 FT DISULFID 551 551
 FT DISULFID 556 556

POTENTIAL.
 THROMBOSPONDIN 1.
 HEPARIN-BINDING (POTENTIAL).
 VWFC.
 TSP TYPE-1 1.
 TSP TYPE-1 2.
 TSP TYPE-1 3.
 EGF-LIKE 1.
 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 3.
 TSP TYPE-3 1.
 TSP TYPE-3 2.
 TSP TYPE-3 3.
 TSP TYPE-3 4.
 TSP TYPE-3 5.
 TSP TYPE-3 6.
 TSP TYPE-3 7.
 C-TERMINAL.
 CELL ATTACHMENT SITE (POTENTIAL).
 INTERCHAIN (PROBABLE).
 BY SIMILARITY.
 BY SIMILARITY.

```
FT DISULFID 575 586 BY SIMILARITY.
FT DISULFID 592 608 BY SIMILARITY.
FT DISULFID 599 617 BY SIMILARITY.
FT DISULFID 620 644 BY SIMILARITY.
FT DISULFID 650 663 BY SIMILARITY.
FT DISULFID 657 676 BY SIMILARITY.
FT DISULFID 678 689 BY SIMILARITY.
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1025 1025 F -> L (IN REF. 2).
SQ SEQUENCE 1170 AA; 129646 MW; 0443E493615E7F06 CRC64;

Query Match 94.3%; Score 50; DB 1; Length 1170;
Best Local Similarity 85.7%; Pred. No. 2.2;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
Db 382 WSPWSEW 388
|||||

RESULT 7
TSPL_XENLA STANDARD; PRT; 1173 AA.
AC P35448;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 1 PRECURSOR.
GN THBS1 OR TSP1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Urry L.A., Ramos J., Duquette M., Desimone D.W., Lawler J.;
RL Submitted (XXX-1993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-
CC V/BETA-3 AND ALPHA-11B/BETA-3 (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 WFEC DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L04278; ; NOT_ANNOTATED_CDS.
CC HSSP; P00740; 11XA.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001881; EGF-Ca.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR003129; TSPN.
CC InterPro; IPR001007; WFEC.
CC InterPro; IPR003367; tsp_3.
CC Pfam; PF000008; EGF; 2.
CC Pfam; PF02210; TSPN; 1.
CC Pfam; PF000090; tsp_1; 3.
```

```
DR Pfam; PF02412; tsp_3; 8.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00001; EGF-like; 1.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00092; TSP1; 3.
DR PROSITE; PS01208; WFEC; 1.
DR Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1173 THROMBOSPONDIN 1.
FT DOMAIN 23 235 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 319 376 WFEC.
FT DOMAIN 382 433 TSP TYPE-1 1.
FT DOMAIN 438 494 TSP TYPE-1 2.
FT DOMAIN 495 546 TSP TYPE-1 3.
FT DOMAIN 550 590 EGF-LIKE 1.
FT DOMAIN 591 648 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 649 693 EGF-LIKE 3.
FT DOMAIN 726 761 TSP TYPE-3 1.
FT DOMAIN 762 784 TSP TYPE-3 2.
FT DOMAIN 785 820 TSP TYPE-3 3.
FT DOMAIN 821 843 TSP TYPE-3 4.
FT DOMAIN 844 881 TSP TYPE-3 5.
FT DOMAIN 882 917 TSP TYPE-3 6.
FT DOMAIN 918 953 TSP TYPE-3 7.
FT DOMAIN 954 1173 C-TERMINAL.
FT SITE 929 931 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 554 565 BY SIMILARITY.
FT DISULFID 559 575 BY SIMILARITY.
FT DISULFID 578 589 BY SIMILARITY.
FT DISULFID 595 611 BY SIMILARITY.
FT DISULFID 602 620 BY SIMILARITY.
FT DISULFID 623 647 BY SIMILARITY.
FT DISULFID 653 666 BY SIMILARITY.
FT DISULFID 660 679 BY SIMILARITY.
FT DISULFID 681 692 BY SIMILARITY.
FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 158 158 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 363 363 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 705 705 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 711 711 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1070 1070 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1173 AA; 130019 MW; A9F036D6516C0F24 CRC64;

Query Match 88.7%; Score 47; DB 1; Length 1173;
Best Local Similarity 71.4%; Pred. No. 6;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
Db 385 WSPWSDW 391
|||||

RESULT 8
TSP2_CHICK
ID TSP2_CHICK STANDARD; PRT; 1178 AA.
AC P35440;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 2 PRECURSOR.
GN THBS2 OR TSP2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
```

```

OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91217026; PubMed=2022631;
RA Lawler J., Duquette M., Ferro P.;
FT "Cloning and sequencing of chicken thrombospondin.";
RL J. Biol. Chem. 266:8039-8043(1991).
CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60853; AAA51437.1; -.
DR PIR; A39804; A39804.
DR HSSP; P00740; 11XA.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; VWFC.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 8.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00001; EGF_Like; 1.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50092; TSP1; 3.
DR PROSITE; PS01208; VWFC; 1.
DR Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1
FT CHAIN ? 1178
FT DOMAIN ? 232
FT DOMAIN 324 381
FT DOMAIN 387 438
FT DOMAIN 441 499
FT DOMAIN 500 553
FT DOMAIN 555 595
FT DOMAIN 596 653
FT DOMAIN 654 698
FT DOMAIN 731 766
FT DOMAIN 767 789
FT DOMAIN 790 825
FT DOMAIN 826 848
FT DOMAIN 849 886
FT DOMAIN 887 922
FT DOMAIN 923 958
FT DOMAIN 959 1178
FT SITE 934 935
FT SITE 559 570
FT DISULFID 564 580
FT DISULFID 583 594
FT DISULFID 600 616

```

```

FT DISULFID 607 625
FT DISULFID 628 652
FT DISULFID 658 671
FT DISULFID 665 684
FT DISULFID 686 697
FT CARBOHYD 157 157
FT CARBOHYD 244 244
FT CARBOHYD 317 317
FT CARBOHYD 322 322
FT CARBOHYD 463 463
FT CARBOHYD 590 590
FT CARBOHYD 716 716
FT CARBOHYD 1075 1075
SQ SEQUENCE 1178 AA; 131816 MW; F37E02F42C87117A2 CRC64;

Query Match      88.7%; Score 47; DB 1; Length 1178;
Best Local Similarity 71.4%; Pred. No. 6;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
Db 390 WSPWSDW 396
||||:|

RESULT 9
PROP_CAVPO STANDARD; PRT; 470 AA.
ID PROP_CAVPO
AC Q64181;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PROPERDIN PRECURSOR.
GN PFC.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Splice;
RX MEDLINE=96148617; PubMed=8550088;
RA Maves K.K., Guenther S.T., Densen P., Moser D.R., Weiler J.M.;
RT "Cloning and characterization of the cDNA encoding guinea-pig
RT properdin: a comparison of properdin from three species.";
RL Immunology 86:475-479(1995).
CC -!- FUNCTION: A POSITIVE REGULATOR OF THE ALTERNATE PATHWAY OF
CC COMPLEMENT. IT BINDS TO AND STABILIZES THE C3-AND C5-CONVERTASE
CC ENZYME COMPLEXES.
CC -!- SIMILARITY: CONTAINS 6 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S81116; AAB35918.1; -.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 6.
DR SMART; SM00209; TSP1; 5.
DR PROSITE; PS50092; TSP1; 6.
KW Signal; Complement alternate pathway; Glycoprotein; Repeat.
FT SIGNAL 1 26
FT CHAIN 27 470
FT DOMAIN 76 133
FT DOMAIN 134 190
FT DOMAIN 191 254
FT DOMAIN 255 312
FT DOMAIN 313 376
FT DOMAIN 379 438
TSP TYPE-1 1.
TSP TYPE-1 2.
TSP TYPE-1 3.
TSP TYPE-1 4.
TSP TYPE-1 5.
TSP TYPE-1 6.

```

FT CARBOHYD 429 429 N-LINKED (GLCNAC...)(POTENTIAL).
SQ SEQUENCE 470 AA; 51431 MW; FDC2B393DC7EC15F CRC64;

Query Match 84.9%; Score 45; DB 1; Length 470;
Best Local Similarity 71.4%; Pred. No. 5.1;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
Db 79 WSPWSSW 85

RESULT 10
VG57_BPMD2 STANDARD; PRT; 152 AA.
AC 084248;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENE 57 PROTEIN (GP57).
GN 57.
OS Mycobacteriophage D29.
OC Viruses.
OX NCBI_TaxID=28369;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98300335; PubMed=9636706;
RA Ford M.E., Sarkis G.J., Belanger A.E., Hendrix R.W., Hatfull G.F.;
RT "Genome structure of mycobacteriophage D29: implications for phage evolution."
RL J. Mol. Biol. 279:143-164(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF022214; AAC18498.1; -
DR InterPro; IPR002936; Toprim.
DR Pfam; PF01751; Toprim; 1.
DR SMART; SM00493; TOPRIM; 1.
SQ SEQUENCE 152 AA; 17194 MW; 75293E268CCE6397 CRC64;

Query Match 81.1%; Score 43; DB 1; Length 152;
Best Local Similarity 71.4%; Pred. No. 3.6;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
Db 13 WSPWRNW 19

RESULT 11
VG57_BPML5 STANDARD; PRT; 152 AA.
AC Q05270;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE GENE 57 PROTEIN (GP57).
GN 57.
OS Mycobacteriophage L5.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
OX NCBI_TaxID=31757;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93211282; PubMed=8459766;
RA Hatfull G.F., Sarkis G.J.;

RT "DNA sequence, structure and gene expression of mycobacteriophage L5:
a phage system for mycobacterial genetics.";
RL Microbiol. 7:395-405(1993).
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z18946; CAA79433.1; -
DR PIR; S31002; S31002.
DR InterPro; IPR002936; Toprim.
DR Pfam; PF01751; Toprim; 1.
DR SMART; SM00493; TOPRIM; 1.
SQ SEQUENCE 152 AA; 17157 MW; 263A969F64DAA69 CRC64;

Query Match 81.1%; Score 43; DB 1; Length 152;
Best Local Similarity 71.4%; Pred. No. 3.6;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
Db 13 WSPWRHW 19

RESULT 12
SSPO_BOVIN STANDARD; PRT; 867 AA.
AC P98167;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SCO-SPONDIN (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Ebendymocyte;
RX MEDLINE=56338614; PubMed=8743952;
RA Gobron S., Monnerie H., Meinzel R., Creveaux I., Lehmann W.,
RA Lamalle D., Dastugue B., Meinzel A.;
RT "SCO-spondin: a new member of the thrombospondin family secreted by
RT the subcommissural organ is a candidate in the modulation of neuronal
RT aggregation.";
RL J. Cell Sci. 109:1053-1061(1996).
CC -; FUNCTION: INVOLVED IN THE MODULATION OF NEURONAL AGGREGATION.
CC -; SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -; TISSUE SPECIFICITY: SUBCOMMISSURAL ORGAN.
CC -; DEVELOPMENTAL STAGE: EMBRYO.
CC -; SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -; SIMILARITY: CONTAINS AT LEAST 4 TSP TYPE-1 DOMAINS.
CC -; SIMILARITY: CONTAINS AT LEAST 2 EGF-LIKE DOMAINS.
CC -; SIMILARITY: CONTAINS AT LEAST 1 F5/8 TYPE C DOMAIN.
CC -; SIMILARITY: CONTAINS AT LEAST 3 LDL-RECEPTOR CLASS A DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X93922; CAA63815.1; -
DR HSSP; P01130; IAJJ.
DR InterPro; IPR000421; FA58_C.

```
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR002919; TIL.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR001007; VWFC.
DR Pfam: PF00754; F5_F8_type_C; 1.
DR Pfam: PF00057; ldl_recept_a; 3.
DR Pfam: PF01826; TIL; 1.
DR Pfam: PF00090; tsp_1; 4.
DR Pfam: PF00093; vwc; 1.
DR SMART: SM00231; FA58C; 1.
DR SMART: SM00192; LDLa; 3.
DR SMART: SM00209; TSP1; 4.
DR SMART: SM00214; VWC; 1.
DR SMART: SM00011; VWC_def; 1.
DR PROSITE: PS01285; FA58C_1; 1.
DR PROSITE: PS01286; FA58C_2; 1.
DR PROSITE: PS01209; LDLRA_1; 3.
DR PROSITE: PS00068; LDLRA_2; 3.
DR PROSITE: PS00092; TSP1; 4.
KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain.
FT NON_TER 1
FT DOMAIN 26 81
FT DOMAIN 103 142
FT DOMAIN 143 180
FT DOMAIN 185 243
FT DOMAIN 344 502
FT DOMAIN 506 544
FT DOMAIN 663 701
FT DOMAIN 723 761
FT DOMAIN 762 813
FT DOMAIN 814 867
FT DISULFID 107 122
FT DISULFID 116 127
FT DISULFID 129 141
FT DISULFID 147 166
FT DISULFID 149 169
FT DISULFID 171 179
FT DISULFID 344 502
FT DISULFID 508 520
FT DISULFID 515 533
FT DISULFID 527 542
FT DISULFID 665 677
FT DISULFID 672 690
FT DISULFID 684 699
FT DISULFID 725 737
FT DISULFID 732 750
FT DISULFID 744 759
FT CARBOHYD 88 88
FT CARBOHYD 309 309
FT CARBOHYD 409 409
FT NON_TER 867 867
SQ SEQUENCE 867 AA; 91817 MW; 9538F2108E787B49 CRC64;

Query Match 79.2%; Score 42; DB 1; Length 867;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSPWAE 6
| | | | |
Db 194 WSPWAE 199

RESULT 13
TRAB_AGRTU
ID TRAB_AGRTU STANDARD; PRT; 421 AA.
AC Q44351;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CONJUGAL TRANSFER PROTEIN TRAB.
GN TRAB.
OS Agrobacterium tumefaciens.

Plasmid pTiC58.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
NCBI_TaxID=358;
[1]
SEQUENCE FROM N.A.
MEDLINE=96312367; PubMed=8763953;
Farrand S.K., Hwang I., Cook D.M.;
"The tra region of the nopaline-type Ti plasmid is a chimera with
elements related to the transfer systems of RSF1010, RP4, and F.";
J. Bacteriol. 178:4233-4247(1996).
CC -!- FUNCTION: ENHANCES CONJUGAL TRANSFER OF THE TI PLASMID.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF010180; AAC17208.1; ..
Conjugation; Transmembrane; Plasmid.
TRANSMEM 33 53
TRANSMEM 103 123 POTENTIAL.
TRANSMEM 144 164 POTENTIAL.
TRANSMEM 169 189 POTENTIAL.
TRANSMEM 193 213 POTENTIAL.
TRANSMEM 287 307 POTENTIAL.
TRANSMEM 366 386 POTENTIAL.
TRANSMEM 388 408 POTENTIAL.
SQ SEQUENCE 421 AA; 45195 MW; 3518583863B0AFC6 CRC64;

Query Match 77.4%; Score 41; DB 1; Length 421;
Best Local Similarity 57.1%; Pred. No. 18;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 WSPWAE 7
| | | | |
Db 327 WQPLQW 333

RESULT 14
CO9_ONCMY
ID CO9_ONCMY STANDARD; PRT; 574 AA.
AC P06682;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COMPLEMENT COMPONENT C9 (FRAGMENT).
GN C9.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OC NCBI_TaxID=8022;
[1]
SEQUENCE FROM N.A.
MEDLINE=88004404; PubMed=2443347;
Stanley K.K., Herz J.;
"Topological mapping of complement component C9 by recombinant DNA
techniques suggests a novel mechanism for its insertion into target
membranes.";
EMBO J. 6:1951-1957(1987).
[2]
REVISONS.
RX MEDLINE=90036879; PubMed=2808363;
Haeffliger J.-A., Tschopp J., Vial N., Jenne D.E.;
"Complete primary structure and functional characterization of the
sixth component of the human complement system. Identification of the
C5b-binding domain in complement C6.";
```

J. Biol. Chem. 264:18041-18051(1989).

CC -!- FUNCTION: C9 IS THE FINAL COMPONENT OF THE COMPLEMENT SYSTEM TO
 CC BE ADDED IN THE ASSEMBLY OF THE MEMBRANE ATTACK COMPLEX. IT IS
 CC ABLE TO ENTER LIPID BILAYERS, FORMING TRANSMEMBRANE CHANNELS.
 CC -!- SIMILARITY: TO COMPLEMENT FACTORS C6, C7, C8, AND TO PERFORIN.
 CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X05474; CAA29037.1; ALT_SEQ.
 DR HSP; B29677; B29677.
 DR PIR; P01130; IAJJ.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002172; LDL_recept_A.
 DR InterPro: IPR001862; MAC_perforin.
 DR InterPro: IPR000884; TSPI.
 DR Pfam: PF00057; ldl_recept_a; 1.
 DR Pfam: PF01823; MACPF; 1.
 DR Pfam: PF00090; tsp_1; 2.
 DR SMART; SM00192; LDLA; 1.
 DR SMART; SM00457; MACPF; 1.
 DR SMART; SM00209; TSPI; 2.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; FALSE_NEG.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS50068; LDLRA_2; 1.
 DR PROSITE; PS00279; MAC_PERFORIN; 1.
 DR PROSITE; PS50092; TSPI; 2.
 KW Complement pathway; Complement alternate pathway; Glycoprotein;
 KW Plasma; Membrane attack complex; Cytolysis; Transmembrane;
 KW EGF-like domain; Repeat.
 FT NON_TER 1 1
 FT DOMAIN 18 76 TSP TYPE-1 1.
 FT DOMAIN 77 117 LDL-RECEPTOR CLASS A.
 FT TRANSMEM 284 300 POTENTIAL.
 FT TRANSMEM 305 324 POTENTIAL.
 FT DOMAIN 475 511 EGF-LIKE.
 FT DOMAIN 528 574 TSP TYPE-1 2.
 FT DISULFID 79 91 BY SIMILARITY.
 FT DISULFID 86 104 BY SIMILARITY.
 FT DISULFID 98 112 BY SIMILARITY.
 FT DISULFID 350 376 BY SIMILARITY.
 FT DISULFID 481 497 BY SIMILARITY.
 FT DISULFID 484 499 BY SIMILARITY.
 FT DISULFID 501 510 BY SIMILARITY.
 FT CARBOHYD 3 3 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 533 533 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 574 AA; 64033 MW; CB516A9F76824D59 CRC64;
 Query Match 77.4%; Score 41; DB 1; Length 574;
 Best Local Similarity 71.4%; Pred. No. 23;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 WSPWAEW 7
 Db 21 WSRWSEW 27
 RESULT 15
 SM5A_HUMAN STANDARD; PRT; 1074 AA.
 ID SM5A_HUMAN
 AC Q13591; O60408;

DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SEMAPHORIN 5A PRECURSOR (SEMAPHORIN F) (SEMA F).
 GN SEMA5A OR SEMAF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98125534; PubMed=9464278;
 RA Simmons A.D., Puschel A.W., McPherson J.D., Overhauser J., Lovett M.;
 RT "Molecular cloning and mapping of human semaphorin F from the Cri-du-
 chat candidate interval.";
 RL Biochem. Biophys. Res. Commun. 242:685-691(1998).
 RN [2]
 RP SEQUENCE OF 1-494 FROM N.A.
 RA Kalicki J., Harmon G.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: MAY ACT AS POSITIVE AXONAL GUIDANCE CUES.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
 CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-1 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U52840; AAC09473.1; --
 DR EMBL: AC004615; AAC14668.1; --
 DR InterPro: IPR003659; PSI.
 DR InterPro: IPR002165; Plexin_repeat.
 DR InterPro: IPR001627; SEMA.
 DR InterPro: IPR000884; TSPI.
 DR Pfam: PF01437; Plexin_repeat; 1.
 DR Pfam: PF01403; SEMA; 1.
 DR Pfam: PF00090; tsp_1; 6.
 DR SMART; SM00423; PSI; 1.
 DR SMART; SM00209; TSPI; 6.
 DR PROSITE; PS50092; TSPI; 5.
 KW Signal; Transmembrane; Repeat; Multigene family; Neurogenesis;
 KW Developmental protein; Glycoprotein.
 FT SIGNAL 1 ?
 FT CHAIN ? 1074 SEMAPHORIN 5A.
 FT DOMAIN ? 968 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 969 989 POTENTIAL.
 FT DOMAIN 990 1074 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 226 507 SEMA.
 FT DOMAIN 540 609 TSP TYPE-1 1.
 FT DOMAIN 595 652 TSP TYPE-1 2.
 FT DOMAIN 653 704 TSP TYPE-1 3.
 FT DOMAIN 707 765 TSP TYPE-1 4.
 FT DOMAIN 784 840 TSP TYPE-1 5.
 FT DOMAIN 841 897 TSP TYPE-1 6.
 FT DOMAIN 898 952 TSP TYPE-1 7.
 FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 367 367 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 717 717 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 933 933 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 56 A -> V (IN REF. 2).

FT CONFLICT 149 149 A -> T (IN REF. 2).
FT CONFLICT 382 382 V -> M (IN REF. 2).
FT CONFLICT 494 494 S -> R (IN REF. 2).
SQ SEQUENCE 1074 AA; 120570 MW; EE3DB763CBE29407 CRC64;

Query Match 77.4%; Score 41; DB 1; Length 1074;
Best Local Similarity 57.1%; Pred. No. 41;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
I:|I I
Db 598 WTPWTSW 604

Search completed: January 31, 2002, 09:36:52
Job time: 771 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:24:52 ; Search time 49.78 seconds
(without alignments)
33.665 Million cell updates/sec

Title: US-09-536-087-8

Perfect score: 128

Sequence: 1 FAENETWVDSCTTCKKFT 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	128	100.0	1172	1 TSHUP2	thrombospondin 2 p
2	128	100.0	1172	2 A42587	thrombospondin 2 p
3	86	67.2	1178	1 A39804	thrombospondin pre
4	68	53.1	1170	1 TSHUP1	thrombospondin 1 p
5	68	53.1	1170	2 A40558	thrombospondin 1 p
6	64	50.0	1822	2 S63985	collagen alpha 2 c
7	63	49.2	1376	2 S63986	collagen alpha 5 c
8	60	46.9	884	2 T18649	hypothetical prote
9	57	44.5	114	2 A34567	beta-microseminopr
10	57	44.5	3198	2 A43426	collagen alpha 2 f
11	54.5	42.6	5376	2 T42215	zonadhesin - mouse
12	54	42.2	114	2 A54663	seminal plasma pro
13	53	41.4	1535	2 S46224	peroxidasin - frui
14	52	40.6	3020	2 A43932	mucin 2 precursor,
15	50.5	39.5	810	2 T10756	Nel-homolog protei
16	49	38.3	837	2 T42112	hypothetical prote
17	49	38.3	1291	2 T21694	hypothetical prote
18	48.5	37.9	395	2 A47701	aspartic proteinas
19	47.5	37.1	835	2 JP0076	nel protein - chic
20	47	36.7	269	2 S36166	paired box transcr
21	47	36.7	364	2 T70880	hypothetical prote
22	47	36.7	2813	1 VWHU	von Willebrand fac
23	47	36.7	3026	2 T28431	variant surface pr
24	47	36.7	4845	2 T31067	BIR repeat contain
25	46	35.9	86	2 A25700	trans-activating t
26	46	35.9	101	2 T09446	tat protein - huma
27	46	35.9	966	1 P1BVBH	RNA la protein - b
28	46	35.9	1042	2 A57534	mucin 5AC (clone L
29	46	35.9	1056	2 A53767	mucin MUC5B, trach

30	46	35.9	2910	2 T42214	otogelin - mouse
31	45.5	35.5	157	2 PN0563	von Willebrand fac
32	45.5	35.5	590	2 E85057	probable transposo
33	45	35.2	373	2 T45043	hypothetical prote
34	45	35.2	392	2 T32524	hypothetical prote
35	45	35.2	489	1 S47076	3-oxoacyl-[acyl-ca
36	45	35.2	977	2 T16232	hypothetical prote
37	45	35.2	1034	2 JC5598	mucin - rat
38	45	35.2	1038	2 T13177	sog protein - frui
39	45	35.2	1131	2 T13177	hypothetical prote
40	45	35.2	1188	2 C71231	hypothetical prote
41	44.5	34.8	199	2 T28981	hypothetical prote
42	44.5	34.8	328	2 C72370	tryptophan--trna 1
43	44.5	34.8	429	2 T21113	hypothetical prote
44	44.5	34.8	531	2 T20763	hypothetical prote
45	44.5	34.8	629	2 T48287	hypothetical prote

ALIGNMENTS

RESULT 1

TSHUP2

thrombospondin 2 precursor - human

C;Species: Homo sapiens (man)

C;Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999

C;Accession: A47379; A42173

R;LaBell, T.L.; Byers, P.H.

Genomics 17, 225-229, 1993

A;Title: Sequence and characterization of the complete human thrombospondin 2 cDNA: p

A;Reference number: A47379; MUID:94010892

A;Accession: A47379

A;Molecule type: mRNA

A;Residues: 1-1172 <LA>

A;Cross-references: GB:LI2350; NID:g307505; PIDN:AAA03703.1; PID:g307506

R;LaBell, T.L.; Milewicz, D.J.; Distche, C.M.; Byers, P.H.

Genomics 12, 421-429, 1992

A;Title: Thrombospondin II: partial cDNA sequence, chromosome location, and expressio

A;Reference number: A42173; MUID:92217961

A;Accession: A42173

A;Molecule type: mRNA

A;Residues: 560-1172 <LA2>

A;Cross-references: GB:M81339

A;Experimental source: fibroblast

A;Note: sequence extracted from NCBI backbone (NCBIN:95091, NCBIP:95096)

C;Genetics:

A;Gene: GDB:THBS2; TSP2

A;Cross-references: GDB:l28789; OMIM:188061

A;Map position: 6q27-6q27

C;Complex: homotrimer, disulfide linked

C;Function:

A;Description: participates in cell migration and adhesion, and in platelet aggregati

C;Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;

C;Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tri

F;1-18/Domain: signal sequence #status predicted <SIG>

F;19-1172/Product: thrombospondin 2 #status predicted <MAT>

F;319-377/Domain: von Willebrand factor type C repeat homology <VWC>

F;380-431/Domain: thrombospondin type 1 repeat homology <THRI>

F;436-492/Domain: thrombospondin type 1 repeat homology <THRI>

F;493-549/Domain: thrombospondin type 1 repeat homology <THRI>

F;553-588/Domain: EGF homology <EGF1>

F;652-691/Domain: EGF homology <EGF>

F;928-930/Region: cell attachment (R-G-D) motif

F;151,316,330,457,584,710,1069/Binding site: carbohydrate (Asn) (covalent) #status pr

F;167-226/Disulfide bonds: #status predicted

F;266,270/Disulfide bonds: interchain #status predicted

F;612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 100.0%; Score 128; DB 1; Length 1172;
Best Local Similarity 100.0%; Pred. No. 3.3e-09;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FAENETWVWVDSCTTCTCKKFKT 22
|||||
DB 327 FAENETWVWVDSCTTCTCKKFKT 348

RESULT 2
A42587

thrombospondin 2 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
C:Accession: A42587; A39851
R:Lawler, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during cell
A:Reference number: A42587; MUID:92147683
A:Accession: A42587
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1172 <LAH>
A:Cross-references: GB:M87275; NID:G340421; PIDN:AAA53064.1; PID:g567241
R:Bornstein, P.; O'Rourke, K.; Wikstrom, K.; Wolf, F.W.; Katz, R.; Li, P.; Dixit, V.M.
J. Biol. Chem. 266, 12821-12824, 1991
A:Title: A second, expressed thrombospondin gene (Thbs2) exists in the mouse genome.
A:Reference number: A39851; MUID:91302287
A:Accession: A39851
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-873 <BOR>
A:Cross-references: GB:M64866; NID:G201994; PIDN:AAA40432.1; PID:g201995
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; von Willebrand factor type C repeat homology
F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>
F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>
F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>
F:553-588/Domain: EGF homology <EGF1>
F:652-691/Domain: EGF homology <EGF>

Query Match 100.08; Score 128; DB 2; Length 1172;
Best Local Similarity 100.08; Pred. No. 3.3e-09;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FAENETWVWVDSCTTCTCKKFKT 22
|||||
DB 327 FAENETWVWVDSCTTCTCKKFKT 348

RESULT 3
A39804

thrombospondin precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A39804
R:Lawler, J.; Duquette, M.; Ferro, P.
J. Biol. Chem. 266, 8039-8043, 1991
A:Title: Cloning and sequencing of chicken thrombospondin.
A:Reference number: A39804; MUID:91217026
A:Accession: A39804
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1178 <LAH>
A:Cross-references: GB:M60853; NID:G212763; PIDN:AAA51437.1; PID:g212764
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; von Willebrand factor type C repeat homology
F:325-383/Domain: von Willebrand factor type C repeat homology <VWC>
F:388-437/Domain: thrombospondin type 1 repeat homology <THR1>
F:442-498/Domain: thrombospondin type 1 repeat homology <THR2>
F:499-555/Domain: thrombospondin type 1 repeat homology <THR3>
F:658-697/Domain: EGF homology <EGF>

Query Match 67.28; Score 86; DB 1; Length 1178;

Best Local Similarity 61.98; Pred. No. 0.00098;
Matches 13; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 FAENETWVWVDSCTTCTCKKFK 21
|||||
DB 333 FAESSEWVDSCTKCTCQDSK 353

RESULT 4
TSHUP1

thrombospondin 1 precursor - human
C:Species: Homo sapiens (man)
C:Date: 23-Aug-1987 #sequence_revision 03-Aug-1995 #text_change 17-Nov-2000
C:Accession: A26155; A34174; A30140; A25812; A05172; A42927
R:Lawler, J.; Hynes, R.O.
J. Cell Biol. 103, 1635-1648, 1986
A:Title: The structure of human thrombospondin, an adhesive glycoprotein with multiple
A:Reference number: A26155; MUID:87057617
A:Accession: A26155
A:Molecule type: mRNA
A:Residues: 1-1170 <LAH>
A:Cross-references: GB:X04665; NID:G37137; PIDN:CAA28370.1; PID:g37138
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Lahterty, C.D.; Glerman, T.M.; Dixit, V.M.
J. Biol. Chem. 264, 11222-11227, 1989
A:Title: Characterization of the promoter region of the human thrombospondin gene. DN
A:Reference number: A34274; MUID:89291870
A:Accession: A34274
A:Molecule type: DNA
A:Residues: 1-166 <LAH>
A:Cross-references: GB:J04835
R:Hennessy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.M.; Rotwei
J. Cell Biol. 108, 729-736, 1989
A:Title: Complete thrombospondin mRNA sequence includes potential regulatory sites in
A:Reference number: A30140; MUID:89139590
A:Accession: A30140
A:Molecule type: mRNA
A:Residues: 1-83, 'A', '85-522, 'A', 524-1170 <HEN>
A:Cross-references: EMBL:X14787; NID:G37464; PIDN:CAA32889.1; PID:g37465
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Kobayashi, S.; Eden-McCutchan, F.; Franson, P.; Bornstein, P.
Biochemistry 25, 8418-8425, 1986
A:Title: Partial amino acid sequence of human thrombospondin as determined by analysis
A:Reference number: A25812; MUID:87157592
A:Accession: A25812
A:Molecule type: mRNA
A:Residues: 1-83, 'A', '85-397 <KOB>
A:Cross-references: GB:M25631; NID:g538353; PIDN:AAA36741.1; PID:g538354
R:Dixit, V.M.; Hennessy, S.W.; Grant, G.A.; Rotwein, P.; Frazier, W.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 5449-5453, 1986
A:Reference number: A05172; MUID:86287276
A:Accession: A05172
A:Molecule type: mRNA
A:Residues: 1-83, 'A', '85-374, 'RC' <DIX>
A:Cross-references: GB:M14326; NID:g340005; PIDN:AAA61237.1; PID:g553801
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Sun, X.; Skorstengaard, K.; Mosher, D.F.
J. Cell Biol. 118, 693-701, 1992
A:Title: Disulfides modulate RGD-inhibitable cell adhesive activity of thrombospondin
A:Reference number: A42927; MUID:92348511
A:Accession: A42927
A:Molecule type: protein
A:Residues: 987-1003 <SUN>
A:Note: Cys-992 is shown to have a free sulfhydryl
C:Genetics:
A:Gene: GDB:THBS1; TSP1; TSP
A:Cross-references: GDB:I20438; OMIM:188060
A:Map position: 15q15-15q15
A:Introns: 23/1
A:Note: the list of introns may be incomplete
C:Complex: homotrimer, disulfide linked
C:Function:
A:Description: participates in cell migration and adhesion, and in platelet aggregati

F;19-1170/Product: thrombospondin 1 #status predicted <MAT>
F;317-375/Domain: von Willebrand factor type C repeat homology <VWC>
F;378-429/Domain: thrombospondin type 1 repeat homology <THR1>
F;434-490/Domain: thrombospondin type 1 repeat homology <THR2>
F;491-547/Domain: thrombospondin type 1 repeat homology <THR3>
F;551-586/Domain: EGF homology <EGF>
F;248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 53.1%; Score 68; DB 2; Length 1170;
Best Local Similarity 50.0%; Pred. No. 0.22;
Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 FAENETWVDSCTTCTCKKFKT 22
: ||| ||||| | |
Db 325 YKNEETVDSCTECHQNSVT 346

RESULT 6
S63985
collagen alpha 2 chain precursor - sea urchin (Strongylocentrotus purpuratus) (fragment)
C:Species: Strongylocentrotus purpuratus (purple urchin)
C:Date: 20-Jul-1996 #sequence_revision 01-Nov-1996 #text_change 23-Aug-1997
C:Accession: S63985; S64637
R:Exposito, J.Y.; Boute, N.; Deleage, G.; Garrone, R.
Eur. J. Biochem. 234, 59-65, 1995
A:Title: Characterization of two genes coding for a similar four-cysteine motif of th
A:Reference number: S63985; MUID:96096722
A:Accession: S63985
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1822 <EXP>
A:Cross-references: EMBL:X89806
R:Exposito, J.Y.
submitted to the EMBL Data Library, July 1995
A:Reference number: S64637
A:Accession: S64637
A:Molecule type: DNA
A:Residues: 1-381, 'DT', 384-677, 'N', 679-1010, 'L', 1012-1822 <EXW>
A:Cross-references: EMBL:X89806
C:Genetics:
A:Gene: COLL2alpha
A:Introns: 33/1; 106/1; 169/2; 254/1; 395/1; 460/3; 534/1; 678/1; 743/3 817/1; 965/1;
C:Superfamily: unassigned collagens; von Willebrand factor type C repeat homology
F;48-106/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 50.0%; Score 64; DB 2; Length 1822;
Best Local Similarity 45.5%; Pred. No. 1;
Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 FAENETWVDSCTTCTCKKFKT 22
: ||| ||||| | |
Db 56 YLHGESWVDECTTCEDNATF 77

RESULT 7
S63986
collagen alpha 5 chain - sea urchin (Strongylocentrotus purpuratus) (fragment)
C:Species: Strongylocentrotus purpuratus (purple urchin)
C:Date: 20-Jul-1996 #sequence_revision 08-Nov-1996 #text_change 25-Apr-1997
C:Accession: S63986; S64638
R:Exposito, J.Y.; Boute, N.; Deleage, G.; Garrone, R.
Eur. J. Biochem. 234, 59-65, 1995
A:Title: Characterization of two genes coding for a similar four-cysteine motif of th
A:Reference number: S63985; MUID:96096722
A:Accession: S63986
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1376 <EXP>
A:Cross-references: EMBL:X89804
R:Exposito, J.Y.
submitted to the EMBL Data Library, July 1995

A:Reference number: S64637
A:Accession: S64638
A:Molecule type: DNA
A:Residues: 1-658, 'G', 660-870, 'G', 872-901, 'H', 903-1185, 'T', 1187-1214, 'Y', 1216-1376 <EXW>
A:Cross-references: EMBL:X89804
C:Genetics:
A:Gene: CDP5alpha
A:Introns: 73/1; 136/2; 221/1; 369/1; 517/1; 659/1; 799/1; 948/1; 1093/1; 1236/1
C:Superfamily: von Willebrand factor type C repeat homology
C:Keywords: extracellular matrix
F:15-73/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 49.2%; Score 63; DB 2; Length 1376;
Best Local Similarity 45.5%; Pred. No. 1.1;
Matches 10; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 1 FAENETWVDSCTTCKKFT 22
DB 23 YLHGEWKVDECTCACDNAT 44

RESULT 8
TI8649
hypothetical protein B0024.14 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: TI8649
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19001
A:Accession: TI8649
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-884 <WIL>
A:Cross-references: EMBL:Z71178; PIDN:CAA94886.1; GSPDB:GN00023; CESP:B0024.14
A:Experimental source: clone B0024
C:Genetics:
A:Gene: CESP:B0024.14
A:Map position: 5
A:Introns: 46/3; 84/1; 212/1; 307/2; 345/2; 394/1; 424/2; 481/1; 596/1; 702/1; 765/3; 85/3

Query Match 46.9%; Score 60; DB 2; Length 884;
Best Local Similarity 44.4%; Pred. No. 1.9;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 FAENETWVDSCTTCTCK 18
DB 519 FTDGETWOLAPCVSCTCR 536

RESULT 9
A34567
beta-microseminoprotein precursor - human
A:Alternate names: beta-inhibin; prostatic secretory protein; PSP-94; seminal plasma protein
C:Species: Homo sapiens (man)
C:Date: 13-Jul-1990 #sequence_revision 13-Jul-1990 #text_change 20-Aug-1999
C:Accession: A34567; A26451; A29777; A30984; A60673; S16238; I52682
R:Green, C.B.; Liu, W.Y.; Kwok, S.C.W.
Biochem. Biophys. Res. Commun. 167, 1184-1190, 1990
A:Title: Cloning and nucleotide sequence analysis of the human beta-microseminoprotein gene
A:Reference number: A34567; MUID:90211299
A:Accession: A34567
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-114 <GR>
A:Cross-references: GB:M34376; NID:g514370; PIDN:AAA59871.1; PID:g514372
R:McKay, M.; Nollet, S.; Fournier, S.; Benjannet, S.; Chappelaine, P.; Paradis, G.; Dubé
DNA 6, 23-29, 1987
A:Title: Molecular cloning and sequence of the cDNA for a 94-amino-acid seminal plasma protein
A:Reference number: A26451; MUID:87161231
A:Accession: A26451

A:Molecule type: mRNA
A:Residues: 1-114 <MBI>
A:Cross-references: GB:M15885; NID:g338414; PIDN:AAA36635.1; PID:g338415
R:Akiyama, K.; Yoshioka, Y.; Schmid, K.; Offner, G.D.; Troxler, R.F.; Tsuda, R.; Hara
Biochim. Biophys. Acta 829, 288-294, 1985
A:Title: The amino acid sequence of human beta-microseminoprotein.
A:Reference number: A29777; MUID:8519974
A:Accession: A29777
A:Molecule type: protein
A:Residues: 21-58, 'PT', 61-113 <AKI>
R:Seidah, N.G.; Arbatti, N.J.; Rochemont, J.; Sheth, A.R.; Chretien, M.
FEBS Lett. 175, 349-355, 1984
A:Title: Complete amino acid sequence of human seminal plasma beta-inhibin. Predictio
A:Reference number: A30984; MUID:85004133
A:Accession: A30984
A:Molecule type: protein
A:Residues: 21-112, 'G', 114 <SEI>
R:Weiber, H.; Andersson, C.; Murne, A.; Rannevik, G.; Lindstroem, C.; Lilja, H.; Fern
Am. J. Pathol. 137, 593-604, 1990
A:Title: Beta microseminoprotein is not a prostate-specific protein.
A:Reference number: A60673; MUID:90379237
A:Accession: A60673
A:Molecule type: protein
A:Residues: 21, 'X', 23-34 <WEI>
A:Experimental source: gastric juice
R:Nollet, S.; McKay, M.; Chretien, M.
Biochim. Biophys. Acta 1089, 247-249, 1991
A:Title: Prostatic secretory protein PSP(94): gene organization and promoter sequence
A:Reference number: S16237; MUID:91274357
A:Accession: S16238
A:Molecule type: DNA
A:Residues: 1-114 <NOL>
A:Cross-references: EMBL:X57928; NID:g35760; PIDN:CAA41002.1; PID:g825707
A:Note: the authors translated the codon ACT for residue 54 as Trp
R:Liu, A.Y.; Bradner, R.C.; Vessella, R.L.
Cancer Lett. 74, 91-99, 1993
A:Title: Decreased expression of prostatic secretory protein PSP94 in prostate cancer
A:Reference number: I52682; MUID:94115955
A:Accession: I52682
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-114 <RES>
A:Cross-references: GB:S67815; NID:g460568; PIDN:AAB29732.1; PID:g460569
C:Comment: This protein is a component of seminal plasma as well as secretory fluids
C:Genetics:
A:Gene: GDB:MSMB
A:Cross-references: GDB:I28042; OMIM:I57145
A:Map position: 10q11.2-10q11.2
A:Introns: 1/3; 37/1; 72/2
C:Superfamily: seminal plasma protein
C:Keywords: semen
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-114/Product: seminal plasma protein #status experimental <MAT>

Query Match 44.5%; Score 57; DB 2; Length 114;
Best Local Similarity 57.1%; Pred. No. 0.99;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 NETWVDSCTTCTC 17
DB 49 NSEWOTDNCETCTC 62

RESULT 10
A43426
collagen alpha 2 fibrillar chain precursor - sea urchin (Strongylocentrotus purpuratu
C:Species: Strongylocentrotus purpuratus (purple urchin)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
C:Accession: A43426
R:Exposito, J.Y.; D'Alessio, M.; Ramirez, F.
J. Biol. Chem. 267, 17404-17408, 1992
A:Title: Novel amino-terminal propeptide configuration in a fibrillar procollagen und

A;Reference number: A43426; MUID:92381062

A;Accession: A43426

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: nucleic acid

A;Residues: 1-3198 <EXP>

A;Cross-references: GB:M92041; NID:g161448; PIDN:AAA30040.1; PID:g161449

A;Note: sequence extracted from NCBI backbone (NCBIP:111365)

C;Superfamily: unassigned collagens; fibrillar collagen carboxyl-terminal homology; von

F;48-106/Domain: von Willebrand factor type C repeat homology <VMC>

F;2978-3198/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 44.5%; Score 57; DB 2; Length 3198;

Best Local Similarity 61.5%; Pred. No. 13;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 ETWVVDSCITCTC 17

I I I I I I I I

Db 60 ESWSVDECTICSC 72

RESULT 11

T42215

zonadhesin - mouse

N;Alternate names: sperm-specific membrane protein

C;Species: Mus musculus (house mouse)

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C;Accession: T42215

R;Gao, Z.; Garbers, D.L.

J. Biol. Chem. 273, 3415-3421, 1998

A;Title: Species diversity in the structure of zonadhesin, a sperm-specific membrane pro

A;Reference number: 222080; MUID:98123114

A;Accession: T42215

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-5376 <GAO>

A;Cross-references: EMBL:U97068; NID:g3327420; PID:g3327421; PIDN:AAC26680.1

C;Genetics:

A;Gene: Zan

A;Map position: 5

C;Function:

A;Description: functions in multiple cell adhesion processes

A;Note: found exclusively on the apical region of the sperm head

C;Keywords: cell adhesion

Query Match 42.6%; Score 54.5; DB 2; Length 5376;

Best Local Similarity 58.8%; Pred. No. 40;

Matches 10; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 3 ENETWVVDSCITCTCK 18

I I I I I I I I

Db 4574 ESKTWVSRGCTKNCTCK 4590

RESULT 12

A54663

seminal plasma protein PSP-94 precursor - rhesus macaque

N;Alternate names: prostatic secretory protein PSP94

C;Species: Macaca mulatta (rhesus macaque)

C;Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 20-Aug-1999

C;Accession: S16237; A54663

R;Nolet, S.; Mbikay, M.; Chretien, M.

Biochim. Biophys. Acta 1089, 247-249, 1991

A;Title: Prostatic secretory protein PSP(94): gene organization and promoter sequence in

A;Reference number: S16237; MUID:91274357

A;Accession: S16237

A;Molecule type: DNA

A;Residues: 1-114 <NO2>

A;Cross-references: EMBL:X57932; NID:g38094; PIDN:CAA41003.1; PID:g829152

A;Note: the authors translated the codon ACT for residue 54 as Trp

R;Nolet, S.; St-Louis, D.; Mbikay, M.; Chretien, M.

Genomics 9, 775-777, 1991

A;Title: Rapid evolution of prostatic protein PSP-94 suggested by sequence divergence

A;Reference number: A54663; MUID:91244325

A;Accession: A54663

A;Molecule type: mRNA

A;Residues: 1-114 <NOL>

A;Cross-references: GB:M92161; NID:g342280; PIDN:AAA36903.1; PID:g342281

C;Genetics:

A;Introns: 1/3; 37/1; 72/2

C;Superfamily: seminal plasma protein

F;1-20/Domain: signal sequence #status predicted <SIG>

Query Match 42.2%; Score 54; DB 2; Length 114;

Best Local Similarity 50.0%; Pred. No. 2.4;

Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 NETWVVDSCITCTCK 19

I I I I I I I I

Db 49 NSKWKTDCERCCTCYK 64

RESULT 13

S46224

peroxidase - fruit fly (Drosophila sp.)

C;Species: Drosophila sp.

C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Oct-2000

C;Accession: S46224

R;Neelson, R.E.; Fessler, L.I.; Takagi, Y.; Blumberg, B.; Keene, D.R.; Olson, P.F.; Pa

EMBO J. 13, 3438-3447, 1994

A;Title: Peroxidase: a novel enzyme-matrix protein of Drosophila development.

A;Reference number: S46224; MUID:94341255

A;Accession: S46224

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1535 <NEL>

A;Cross-references: GB:U11052; NID:g531384; PIDN:AAA61568.1; PID:g531385

C;Superfamily: peroxidase; myeloperoxidase homology; proteoglycan amino-terminal hom

F;19-44/Domain: proteoglycan amino-terminal homology <PAHA>

F;661-1350/Domain: myeloperoxidase homology <MPX>

Query Match 41.4%; Score 53; DB 2; Length 1535;

Best Local Similarity 57.1%; Pred. No. 24;

Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 NETWVVDSCITCTC 17

I I I I I I I I

Db 1478 NEVWSPDVCTKCNC 1491

RESULT 14

A43932

mucin 2 precursor, intestinal - human (fragments)

N;Alternate names: mucin SMUC-41

C;Species: Homo sapiens (man)

C;Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999

C;Accession: A49963; A45106; B45106; A45932; B33532; A61257; PQ0328; PQ0329

R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.

J. Biol. Chem. 269, 2440-2446, 1994

A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t

A;Reference number: A49963; MUID:94132002

A;Accession: A49963

A;Molecule type: mRNA

A;Residues: 1-639 <GU1>

A;Cross-references: GB:I21998

R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.

J. Biol. Chem. 267, 21375-21383, 1992

A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up

A;Reference number: A45106; MUID:93016075

A;Accession: A45106

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 626-1895 <GU2>

A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396
 A:Note: sequence extracted from NCBI backbone (NCBIP:116706)
 A:Accession: B45106
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 2037-3020 <GU3>
 A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398
 A:Experimental source: colon
 A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
 R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.
 J. Clin. Invest. 88, 1005-1013, 1991
 A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorphisms
 A:Reference number: A43932; MUID:91358717
 A:Accession: A43932
 A:Molecule type: DNA
 A:Residues: 1343-1350
 A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
 A:Note: sequence inconsistent with the nucleotide translation
 A:Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)
 R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.
 J. Biol. Chem. 264, 6480-6487, 1989
 A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evidence
 A:Reference number: A33532; MUID:89197956
 A:Accession: B33532
 A:Molecule type: mRNA
 A:Residues: 1916-2193 <GU4>
 A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
 A:Experimental source: intestine
 R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
 J. Clin. Invest. 87, 77-82, 1991
 A:Title: Human bronchus and intestine express the same mucin gene.
 A:Reference number: A61257; MUID:91086481
 A:Accession: A61257
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
 A:Experimental source: bronchus
 R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner,
 Biochem. Biophys. Res. Commun. 183, 821-828, 1992
 A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-
 A:Reference number: PQ0328; MUID:92198477
 A:Accession: PQ0328
 A:Molecule type: mRNA
 A:Residues: 2328-2468 <XUG>
 A:Cross-references: GB:M86523
 A:Experimental source: small intestine
 A:Accession: PQ0329
 A:Molecule type: protein
 A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
 C:Genetics:
 A:Gene: GDB:MUC2
 A:Cross-references: GDB:120203; OMIM:158370
 A:Map position: 11p15.5-11p15.5
 C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von
 C:Keywords: glycoprotein; intestine; tandem repeat
 F:2766-2834/Domain: von Willebrand factor type C repeat homology <VMC>

Query Match 40.6%; Score 52; DB 2; Length 3020;
 Best Local Similarity 56.2%; Pred. No. 54;
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 ENETWVDSCTTCTCK 18
 ||||| : | |||
 DB 2262 ENETWILDCDFWATCK 2277

RESULT 15

T10756

Nel-homolog protein - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999

C:Accession: T10756

R:Kuroda, S.; Tokunaga, C.; Kiyohara, Y.; Konishi, H.; Matsuhashi, S.; Kikkawa, U.
 submitted to the EMBL Data Library, November 1998
 A:Description: Protein Kinase C-binding protein.
 A:Reference number: Z17122

A:Accession: T10756

A:Status: Preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-810 <KUR>

A:Cross-references: EMBL:U48246; NID:g3851179; PID:g3851180

A:Experimental source: strain Sprague-Dawley, brain

Query Match 39.5%; Score 50.5; DB 2; Length 810;
 Best Local Similarity 42.1%; Pred. No. 31;
 Matches 8; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

QY 1 FAENETWV-VDSCTTCTCK 18

DB 280 YRDQDSWVDGDCGNCCTCK 298

Search completed: January 31, 2002, 09:24:54
 Job time: 259 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:24:54 ; Search time 49.78 seconds
(without alignments)
22.953 Million cell updates/sec

Title: US-09-536-087-9
Perfect score: 82
Sequence: 1 ELIGPPKTRNMSAC 15
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	1172	1 TSHUP2	thrombospondin 2 p
2	72	87.8	1172	2 A42587	thrombospondin 2 p
3	44	53.7	614	1 W1W41	El protein - human
4	42	51.2	432	2 S38993	uracil transport p
5	40.5	49.4	958	2 T26258	hypothetical prote
6	40	48.8	514	2 T48987	hypothetical prote
7	40	48.8	673	2 A45456	NADH dehydrogenase
8	39	47.6	81	2 B70966	hypothetical prote
9	39	47.6	439	1 DEPSHA	homoserine dehydro
10	39	47.6	604	2 S36493	El protein - human
11	39	47.6	605	2 S36469	El protein - human
12	39	47.6	651	2 A96781	unknown protein F9
13	39	47.6	731	2 T15664	hypothetical prote
14	39	47.6	2437	2 S42612	transmembrane prot
15	38.5	47.0	391	2 T34935	hypothetical prote
16	38	46.3	207	2 T38905	ATP-binding cassat
17	38	46.3	209	2 T30242	cystathione syntha
18	38	46.3	260	2 C83362	hypothetical prote
19	38	46.3	363	2 T23609	hypothetical prote
20	38	46.3	414	2 D64111	uracil transport p
21	38	46.3	434	2 A57986	membrane-bound ura
22	38	46.3	470	2 D69012	conserved hypotet
23	38	46.3	552	2 E72283	alpha-galactosidas
24	38	46.3	556	2 G86319	hypothetical prote
25	38	46.3	1024	2 T16491	hypothetical prote
26	38	46.3	1964	2 T09059	notch4 - mouse
27	37.5	45.7	568	2 E96648	hypothetical prote
28	37	45.1	128	2 C86179	hypothetical prote
29	37	45.1	159	2 A85022	probable actin pol

30	37	45.1	177	2 C37141	parB protein - Esc
31	37	45.1	295	2 F83139	geranyltransferasf
32	37	45.1	298	2 T09145	ethylene-forming e
33	37	45.1	348	2 D69009	hydrogenase expres
34	37	45.1	368	2 T03580	probable transcrip
35	37	45.1	403	2 D81844	uracil permease NM
36	37	45.1	403	2 B81096	uracil permease NM
37	37	45.1	411	2 C70849	hypothetical prote
38	37	45.1	440	2 I54442	peroxisome prolife
39	37	45.1	440	2 JC4530	peroxisome prolife
40	37	45.1	465	2 S33680	ribonucleoprotein
41	37	45.1	471	2 S33679	ribonucleoprotein
42	37	45.1	494	2 JC4399	monocarboxylate tr
43	37	45.1	545	1 COBYC2	cyclin 2 - yeast (
44	37	45.1	604	2 S36540	El protein - human
45	37	45.1	604	2 S36487	El protein - human

ALIGNMENTS

RESULT 1
TSHUP2
thrombospondin 2 precursor - human
C:Species: Homo sapiens (man)
C:Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999
C:Accession: A47379; A42173
R:Labell, T.L.; Byers, P.H.
Genomics 17, 225-229, 1993
A:Title: Sequence and characterization of the complete human thrombospondin 2 cDNA: p
A:Reference number: A47379; MUID:94010892
A:Accession: A47379
A:Molecule type: mRNA
A:Residues: 1-1172 <LAB>
A:Cross-references: GB:L12350; NID:g307505; PIDN:AAA03703.1; PID:g307506
R:Labell, T.L.; Milewicz, D.J.; Distche, C.M.; Byers, P.H.
Genomics 12, 421-429, 1992
A:Title: Thrombospondin II: partial cDNA sequence, chromosome location, and expressio
A:Reference number: A42173; MUID:92217961
A:Accession: A42173
A:Molecule type: mRNA
A:Residues: 560-1172 <LA2>
A:Cross-references: GB:M81339
A:Experimental source: fibroblast
A:Note: sequence extracted from NCBI backbone (NCBIN:95091, NCBIP:95096)
C:Genetics:
A:Gene: GDB:THBS2; TSP2
A:Cross-references: GDB:L28789; OMIM:188061
A:Map position: 6q27-6q27
C:Complex: homotrimer, disulfide linked
C:Function:
A:Description: participates in cell migration and adhesion, and in platelet aggregati
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;
C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tri
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-1172/Product: thrombospondin 2 #status predicted <MAT>
F:319-377/Domain: von willebrand factor type C repeat homology <VWC>
F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>
F:493-548/Domain: thrombospondin type 1 repeat homology <THR3>
F:553-588/Domain: EGF homology <EGF1>
F:652-691/Domain: EGF homology <EGF>
F:928-930/Region: cell attachment (R-G-D) motif
F:151,316,330,457,584,710,1069/Binding site: carbohydrate (Asn) (covalent) #status pr
F:167-226/Disulfide bond: #status predicted
F:266,270/Disulfide bonds: interchain #status predicted
F:612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 100.0%; Score 82; DB 1; Length 1172;
Best Local Similarity 100.0%; Pred. NO. 4.le-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ELIGGPPKTRNMSAC 15
 :: |||||
 Db 306 ELIGGPPKTRNMSAC 320

RESULT 2
 A42587
 thrombospondin 2 precursor - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
 C:Accession: A42587; A39851
 R:Laberty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
 J. Biol. Chem. 267, 3274-3281, 1992
 A>Title: Characterization of mouse thrombospondin 2 sequence and expression during cell
 A:Reference number: A42587; MUID:92147683
 A:Accession: A42587
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-1172 <LAH>
 A:Cross-references: GB:L07803; GB:M87275; NID:g340421; PIDN:AAA53064.1; PID:g567241
 A>Note: sequence extracted from NCBI backbone (NCBIP:81502)
 R:Bornstein, P.; O'Rourke, K.; Wikstrom, K.; Wolf, F.W.; Katz, R.; Li, P.; Dixit, V.M.
 J. Biol. Chem. 266, 12821-12824, 1991
 A>Title: A second, expressed thrombospondin gene (Thbs2) exists in the mouse genome.
 A:Reference number: A39851; MUID:91302287
 A:Accession: A39851
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-873 <BOR>
 A:Cross-references: GB:M64866; NID:g201994; PIDN:AAA0432.1; PID:g201995
 C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; von
 C:Keywords: calcium binding; glycoprotein
 F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>
 F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
 F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>
 F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>
 F:553-588/Domain: EGF homology <EGF1>
 F:652-691/Domain: EGF homology <EGF>

Query Match 87.8%; Score 72; DB 2; Length 1172;
 Best Local Similarity 93.3%; Pred. NO. 0.00023;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ELIGGPPKTRNMSAC 15
 :: |||||
 Db 306 ELIGGPPKTRNMSAC 320

RESULT 3
 W1WL41
 E1 protein - human papillomavirus type 41
 C:Species: human papillomavirus type 41
 A>Note: host Homo sapiens (man)
 C>Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jul-1999
 C:Accession: C43550
 R:Hirt, L.; Hirsch-Behnman, A.; De Villiers, E.M.
 Virus Res. 18, 179-190, 1990
 A>Title: Nucleotide sequence of human papillomavirus (HPV) type 41: an unusual HPV type
 A:Reference number: A43550
 A:Accession: C43550
 A>Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-614 <HIP>
 A:Cross-references: EMBL:X56147; NID:g60942; PIDN:CAA39614.1; PID:g60945
 C:Superfamily: papillomavirus E1 protein
 C:Keywords: early protein

Query Match 53.7%; Score 44; DB 1; Length 614;
 Best Local Similarity 53.8%; Pred. NO. 9.5;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 LIGGPPKTRNMSA 14
 :: |||||
 Db 270 MLANPPTRNMPA 282

RESULT 4
 S38893
 uracil transport protein - Bacillus caldolyticus
 N:Alternate names: uracil permease
 C:Species: Bacillus caldolyticus
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-Sep-1999
 C:Accession: S38893
 R:Ghim, S.Y.; Neuhard, J.
 submitted to the EMBL Data Library, November 1993
 A>Description: The pyrimidine biosynthesis operon of the thermophile Bacillus caldoly
 A:Reference number: S38892
 A:Accession: S38893
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-432 <GHI>
 A:Cross-references: EMBL:X76083; NID:g431229; PIDN:CAA53697.1; PID:g431231
 C:Superfamily: uracil transport protein uraA
 C:Keywords: transmembrane protein

Query Match 51.2%; Score 42; DB 2; Length 432;
 Best Local Similarity 87.5%; Pred. NO. 15;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LIGGPPKPT 9
 :: |||||
 Db 284 LLGGPPKPT 291

RESULT 5
 T26258
 hypothetical protein W07A8.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T26258
 R:Hasham, V.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z20184
 A:Accession: T26258
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-958 <WIL>
 A:Cross-references: EMBL:Z82075; PIDN:CAB04929.1; GSPDB:GN000023; CESP:W07A8.3
 A:Experimental source: clone W07A8
 C:Genetics:

A:Gene: CESP:W07A8.3
 A:Map position: 5
 A:Introns: 17/2; 35/3; 78/3; 133/2; 167/3; 188/1; 244/3; 359/3; 403/2; 436/1; 499/3;
 Query Match 49.4%; Score 40.5; DB 2; Length 958;
 Best Local Similarity 40.0%; Pred. NO. 62;
 Matches 10; Conservative 2; Mismatches 2; Indels 11; Gaps 1;

QY 1 ELIGG-----PPKTRNMSA 14
 :: ||||
 Db 700 DLGGFSPMKPQTSTPTNTPNSA 724

RESULT 6
 T48987
 hypothetical protein F25L23.50 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Dec-2000
 C:Accession: T48987
 R:D'Angelio, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.
 submitted to the Protein Sequence Database, May 2000
 A:Reference number: Z25012

A:Accession: T48987
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1514 <DNA>
 A:Cross-references: EMBL:AL356014; GSPDB:GN00061; ATSP:F25L23.50
 A:Experimental source: cultivar Columbia; BAC clone F25L23
 C:Genetics:
 A:Gene: ATSP:F25L23.50
 A:Map position: 3
 A:Introns: 317/3: 365/3
 C:Superfamily: Arabidopsis thaliana hypothetical protein FL7JL6.30

Query Match 48.8%; Score 40; DB 2; Length 514;
 Best Local Similarity 58.3%; Pred. No. 39;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 GPPPKTRMSAC 15
 DB 179 GPPPKTRMSAC 190
 |||| :|||

RESULT 7
 A45456
 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain NQ03 - Paracoccus denitrificans
 C:Species: Paracoccus denitrificans
 C:Date: 24-Feb-1994 #sequence_revision 12-Apr-1996 #text_change 11-Jun-1999
 C:Accession: S23948; A45456
 R:Xu, X.; Matsuno-Yagi, A.; Yagi, T.
 Arch. Biochem. Biophys. 296, 40-48, 1992
 A:Title: Structural features of the 66-kDa subunit of the energy-transducing NADH-ubiquinone reductase
 A:Reference number: S23946; MUID:92296779
 A:Accession: S23948
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-673 <XUA>
 A:Cross-references: GB:M84572; NID:g150599; PIDN:AAA25587.1; PID:g150601
 R:Xu, X.; Matsuno-Yagi, A.; Yagi, T.
 Biochemistry 32, 968-981, 1993
 A:Title: DNA sequencing of the seven remaining structural genes of the gene cluster encoding NADH dehydrogenase
 A:Reference number: A45456; MUID:93136200
 A:Accession: A45456
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 658-673 <XU1>
 A:Note: sequence extracted from NCBI backbone (NCBI:123409, NCBI:123410)
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 11
 C:Keywords: NAD; oxidoreductase

Query Match 48.8%; Score 40; DB 2; Length 673;
 Best Local Similarity 75.0%; Pred. No. 52;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ELIGGPPK 8
 DB 54 EVVGGPPK 61
 ||:||||

RESULT 8
 B70966
 hypothetical protein Rv2654c - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: B70966
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987
 A:Accession: B70966

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-81 <COL>
 A:Cross-references: GB:Z80225; GB:AL123456; NID:g3242265; PIDN:CAB02330.1; PID:e26640
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: Rv2654c

Query Match 47.6%; Score 39; DB 2; Length 81;
 Best Local Similarity 50.0%; Pred. No. 8.6;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 ELIGGPPKTRNMSA 14
 DB 17 ELVGGPPVEASAAA 30
 ||:|||| :||

RESULT 9
 DEPSHA
 homoserine dehydrogenase (EC 1.1.1.3) - Pseudomonas aeruginosa
 C:Species: Pseudomonas aeruginosa
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 11-Jun-1999
 C:Accession: S27979; S19918
 R:Clepet, C.; Borne, F.; Krishnapillai, V.; Baird, C.; Patte, J.C.; Cami, B.
 Mol. Microbiol. 6, 3109-3119, 1992
 A:Title: Isolation, organization and expression of the Pseudomonas aeruginosa threonine dehydratase gene
 A:Reference number: S27979; MUID:93086420
 A:Accession: S27979
 A:Molecule type: DNA
 A:Residues: 1-439 <CIE>
 A:Cross-references: EMBL:X65033; NID:g45328; PIDN:CAA46167.1; PID:g45329
 C:Genetics:
 A:Gene: hom

A:Map position: 31 min
 C:Superfamily: homoserine dehydrogenase; homoserine dehydrogenase homology
 C:Keywords: NADP; oxidoreductase; threonine biosynthesis
 F:6-246/Domain: homoserine dehydrogenase homology <HSD>

Query Match 47.6%; Score 39; DB 1; Length 439;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ELIGGPP 7
 DB 81 ELIGGPP 87
 |||||

RESULT 10
 S36493
 El protein - human papillomavirus type 25
 C:Species: human papillomavirus type 25
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999
 C:Accession: S36493
 R:Deilius, H.; Hofmann, B.
 submitted to the EMBL Data Library, August 1993
 A:Description: Primer-directed sequencing of human papillomavirus types.
 A:Reference number: S36469
 A:Accession: S36493
 A:Molecule type: DNA
 A:Residues: 1-604
 A:Cross-references: EMBL:X74471; NID:g396948; PIDN:CAA52526.1; PID:g396951
 C:Superfamily: papillomavirus El protein
 C:Keywords: early protein; nucleus

Query Match 47.6%; Score 39; DB 2; Length 604;
 Best Local Similarity 42.9%; Pred. No. 70;
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 ELIGGPPKTRNMSA 14
 :||| |||:

Db 258 QILSEPPKLRNVAA 271

RESULT 11

S36469

El protein - human papillomavirus type 14D

C:Species: human papillomavirus type 14D

C:Date: 09-Dec-1993 #sequence_revision 26-Jul-1996 #text_change 26-Aug-1999

C:Accession: S36469

R:Delius, H.; Hofmann, B.

A:Description: Primer-directed sequencing of human papillomavirus types.

A:Reference number: S36469

A:Accession: S36469

A:Molecule type: DNA

A:Residues: 1-605

A:Cross-references: EMBL:X74467; NID:g396918; PIDN:CAA52502.1; PID:g396919

A:Experimental source: strain 14D

C:Superfamily: papillomavirus El protein

C:Keywords: early protein; nucleus

Query Match 47.6%; Score 39; DB 2; Length 605;

Best Local Similarity 42.9%; Pred. No. 70;

Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 ELIGGPPKTRNSA 14

::: ||| ||::|

Db 259 QILSEPPKLRNVAA 272

RESULT 12

A96781

unknown protein F9E10.5 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: A96781

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: A96781

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-651 <STO>

A:Cross-references: GB:AE005173; NID:g6646755; PIDN:AAF21067.1; GSPDB:GN00141

C:Genetics:

A:Gene: F9E10.5

A:Map position: 1

Query Match 47.6%; Score 39; DB 2; Length 651;

Best Local Similarity 54.5%; Pred. No. 76;

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1.ELIGGPPKTRN 11

::: ||||| :

Db 35 DVFGGPPKRRS 45

RESULT 13

T15664

hypothetical protein C27F2.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T15664

R:Kirsten, J.

A:Submitted to the EMBL Data Library, November 1995

A:Description: The sequence of C. elegans cosmid c27F2.

A:Reference number: Z18385

A:Accession: T15664

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-731 <KIR>

A:Cross-references: EMBL:U40419; NID:g1065504; PID:g1065510; PIDN:AAA81427.1; CESP:C2

C:Genetics:

A:Gene: CESP:C27F2.7

A:Introns: 64/3; 162/3; 211/1; 270/1; 311/2; 352/3; 374/1; 398/2; 423/1; 467/1; 495/3

Query Match 47.6%; Score 39; DB 2; Length 731;

Best Local Similarity 64.3%; Pred. No. 85;

Matches 9; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

QY 4 GGPP--KTRNMSAC 15

| | | | | | |

Db 389 GKPPPVKTRNTSRC 402

RESULT 14

S42612

transmembrane protein precursor - zebra fish

C:Species: Brachydanio rerio (zebra fish)

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Sep-1999

C:Accession: S42612

R:Bierkamp, C.; Campos-Ortega, J.A.

Mech. Dev. 43, 87-100, 1993

A:Title: A zebrafish homologue of the Drosophila neurogenic gene Notch and its patter

A:Reference number: S42612; MUID:94128602

A:Accession: S42612

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2437 <BIE>

A:Cross-references: EMBL:X69088; NID:g433866; PIDN:CAA48831.1; PID:g433867

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol

F:755-786/Domain: EGF homology <EGF1>

F:1023-1054/Domain: EGF homology <EGF>

F:1185-1216/Domain: EGF homology <EGF2>

F:1915-1947/Domain: ankyrin repeat homology <AN1>

F:1948-1980/Domain: ankyrin repeat homology <AN2>

F:1982-2014/Domain: ankyrin repeat homology <AN3>

F:2015-2047/Domain: ankyrin repeat homology <AN4>

F:2048-2080/Domain: ankyrin repeat homology <AN5>

Query Match 47.6%; Score 39; DB 2; Length 2437;

Best Local Similarity 77.8%; Pred. No. 3e+02;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GGPPKTRNM 12

| | | | | | |

Db 1717 GGPKTGEM 1725

RESULT 15

T34935

hypothetical protein SC3F9.13 SC3F9.13 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 08-Sep-2000

C:Accession: T34935

R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

A:Submitted to the EMBL Data Library, June 1998

A:Reference number: Z21562

A:Accession: T34935

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-391 <SEE>

A:Cross-references: EMBL:AL023862; PIDN:CAA19636.1; GSPDB:GN00070; SCOEDB:SC3F9.13

A:Experimental source: strain A3(2)

C;Genetics:
A;Gene: SCOEDB:SC3F9.13
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE29.07c

Query Match 47.0%; Score 38.5; DB 2; Length 391;
Best Local Similarity 66.7%; Pred. No. 54;
Matches 8; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

```
QY      2  LIGGP-PKTRNM 12
          | : | | | | | | :
Db     334  LVGGPHPLTRNL 345
```

Search completed: January 31, 2002, 09:24:55
Job time: 260 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:24:55 ; Search time 49.78 Seconds
(without alignments)
10.712 Million cell updates/sec

Title: US-09-536-087-10

Perfect score: 53

Sequence: 1 WSPWAEW 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	100.0	1172	1 TSHUP2	thrombospondin 2 p
2	53	100.0	1172	2 A42587	thrombospondin 2 p
3	50	94.3	802	2 T21315	hypothetical prote
4	50	94.3	1170	1 TSHUP1	thrombospondin 1 p
5	50	94.3	1170	2 A40558	thrombospondin 1 p
6	48	90.6	788	2 T25061	hypothetical prote
7	47	88.7	1178	1 A39804	thrombospondin pre
8	44	83.0	1184	2 T09484	cartilage intermed
9	43	81.1	152	2 S31002	gene 57 protein -
10	43	81.1	152	2 G72806	gp57 protein - Myc
11	42	79.2	213	2 T44588	probable transcrip
12	42	79.2	654	2 T29247	hypothetical prote
13	41	77.4	421	2 T03421	traB protein - Agr
14	41	77.4	574	2 B29677	complement C9 prec
15	41	77.4	708	2 G96518	protein T2E6.8 [im
16	41	77.4	736	2 T19366	hypothetical prote
17	41	77.4	957	2 T15976	hypothetical prote
18	41	77.4	1074	2 JCS928	semaphorin F precu
19	41	77.4	1444	2 T18956	angiogenesis inhib
20	40	75.5	169	2 C75595	hypothetical prote
21	40	75.5	387	2 T02780	probable conjugal
22	40	75.5	702	2 B85091	isoamylase-like pr
23	40	75.5	990	1 C46335	env polypeptide pr
24	39	73.6	142	2 C75271	hypothetical prote
25	39	73.6	204	2 T15295	hypothetical prote
26	39	73.6	320	2 F64909	hypothetical prote
27	39	73.6	335	2 T39425	hypothetical prote
28	39	73.6	383	2 T30519	dihydroorotase (EC
29	39	73.6	515	2 S67290	probable membrane

complement C9 prec
mucin JER57 - huma
F-spondin precursor
F-spondin - rat
hypothetical prote
unc-5 protein - Ca
gene ADAMTS-1 prot
lysozyme (EC 3-2-1
hypothetical prote
conserved hypotnet
glyceroldehyde-3-p
N-formyl peptide r
ferredoxin--NADP+
hypothetical prote
probable isoamylas

ALIGNMENTS

RESULT 1

TSHUP2

thrombospondin 2 precursor - human

C:Species: Homo sapiens (man)

C>Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999

C:Accession: A47379; A42173

R:LaBell, T.L.; Byers, P.H.

Genomics 17, 225-229, 1993

A:Title: Sequence and characterization of the complete human thrombospondin 2 cDNA: p

A:Reference number: A47379; MUID:94010892

A:Accession: A47379

A:Molecule type: mRNA

A:Residues: 1-1172 <LAB>

A:Cross-references: GB:L12350; NID:g307505; PIDN:AAA03703.1; PID:g307506

R:LaBell, T.L.; Milewicz, D.J.; Distech, C.M.; Byers, P.H.

Genomics 12, 421-429, 1992

A:Title: Thrombospondin II: partial cDNA sequence, chromosome location, and expressio

A:Reference number: A42173; MUID:92217961

A:Accession: A42173

A:Molecule type: mRNA

A:Residues: 560-1172 <LA2>

A:Cross-references: GB:M81339

A:Experimental source: fibroblast

A:Note: sequence extracted from NCBI backbone (NCBIN:95091, NCBI:P:95096)

C:Genetics:

A:Gene: GDB:THBS2; TSP2

A:Cross-references: GDB:128789; OMIM:188061

A:Map position: 6q27-6q27

C:Complex: homotrimer, disulfide linked

C:Function:

A:Description: participates in cell migration and adhesion, and in platelet aggregati

C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;

C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tri-

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-1172/Product: thrombospondin 2 #status predicted <MAT>

F:319-377/Domain: von Willebrand factor type C repeat homology <WVC>

F:380-431/Domain: thrombospondin type 1 repeat homology <THRI>

F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>

F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>

F:553-588/Domain: EGF homology <EGF1>

F:652-691/Domain: EGF homology <EGF>

F:928-930/Region: cell attachment (R-G-D) motif

F:151,316,330,457,584,710,1069/Binding site: carbohydrate (Asn) (covalent) #status pr

F:167-226/Disulfide bonds: #status predicted

F:266,270/Disulfide bonds: interchain #status predicted

F:612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 100.0%; Score 53; DB 1; Length 1172;

Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
|||||
Db 384 WSPWAEW 390

RESULT 2

A42587
thrombospondin 2 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
C:Accession: A42587; A39851
R:Lahter, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during cell
A:Reference number: A42587; MUID:92147683
A:Accession: A42587
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1172 <LAH>
A:Cross-references: GB:M87275; NID:G340421; PIDN:AAA53064.1; PID:G567241
A:Note: sequence extracted from NCBI backbone (NCBIP:81502)
R:Bornstein, P.; O'Rourke, K.; Wikstrom, K.; Wolf, F.W.; Katz, R.; Li, P.; Dixit, V.M.
J. Biol. Chem. 266, 12821-12824, 1991
A:Title: A second, expressed thrombospondin gene (Thbs2) exists in the mouse genome.
A:Reference number: A39851; MUID:91302287
A:Accession: A39851
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-873 <BOR>
A:Cross-references: GB:M64866; NID:G201994; PIDN:AAA0432.1; PID:G201995
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; v
C:Keywords: calcium binding; glycoprotein
F:319-377/Domain: von Willebrand factor type C repeat homology <YWC>
F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>
F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>
F:553-588/Domain: EGF homology <EGF1>
F:652-691/Domain: EGF homology <EGF>

Query Match 100.0%; Score 53; DB 2; Length 1172;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
|||||
Db 384 WSPWAEW 390

RESULT 3

T21315
hypothetical protein F23H12.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21315
R:Kershaw, J.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19405
A:Accession: T21315
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-802 <WIL>
A:Cross-references: EMBL:Z74472; PIDN:CAA98943.1; GSPDB:GN00023; CESP:F23H12.5
A:Experimental source: clone F23H12
C:Genetics:
A:Gene: CESP:F23H12.5
A:Map position: 5
A:Introns: 24/1; 64/3; 109/1; 133/1; 522/3; 548/3; 663/3; 704/1; 740/1

Query Match 94.3%; Score 50; DB 2; Length 802;
Best Local Similarity 85.7%; Pred. No. 3.4;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
|||||
Db 706 WSPWAEW 712

RESULT 4

TSHUP1
thrombospondin 1 precursor - human
C:Species: Homo sapiens (man)
C:Date: 23-Aug-1987 #sequence_revision 03-Aug-1995 #text_change 17-Nov-2000
C:Accession: A26155; A34274; A30140; A25812; A05172; A42927
R:Lawler, J.; Hynes, R.O.
J. Cell Biol. 103, 1635-1648, 1986
A:Title: The structure of human thrombospondin, an adhesive glycoprotein with multipl
A:Reference number: A26155; MUID:87057617
A:Accession: A26155
A:Molecule type: mRNA
A:Residues: 1-1170 <LAH>
A:Cross-references: GB:X04665; NID:G37137; PIDN:CAA28370.1; PID:G37138
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Lahter, C.D.; Gierman, T.M.; Dixit, V.M.
J. Biol. Chem. 264, 11222-11227, 1989
A:Title: Characterization of the promoter region of the human thrombospondin gene. DN
A:Reference number: A34274; MUID:89291870
A:Accession: A34274
A:Molecule type: DNA
A:Residues: 1-166 <LAH>
A:Cross-references: GB:J04835
R:Hennesy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.M.; Rotwei
J. Cell Biol. 108, 729-736, 1989
A:Title: Complete thrombospondin mRNA sequence includes potential regulatory sites in
A:Reference number: A30140; MUID:89139590
A:Accession: A30140
A:Molecule type: mRNA
A:Residues: 1-83, 'A', '85-522, 'A', 524-1170 <HEN>
A:Cross-references: EMBL:X14787; NID:G37464; PIDN:CAA32889.1; PID:G37465
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Kobayashi, S.; Eden-McCutchan, F.; Framson, P.; Bornstein, P.
Biochemistry 25, 8418-8425, 1986
A:Title: Partial amino acid sequence of human thrombospondin as determined by analysi
A:Reference number: A25812; MUID:87157592
A:Accession: A25812
A:Molecule type: mRNA
A:Residues: 1-83, 'A', '85-397 <KOB>
A:Cross-references: GB:M25631; NID:G538353; PIDN:AAA36741.1; PID:G538354
R:Dixit, V.M.; Hennesy, S.W.; Grant, G.A.; Rotwein, P.; Frazier, W.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 5449-5453, 1986
A:Reference number: A05172; MUID:86287276
A:Accession: A05172
A:Molecule type: mRNA
A:Residues: 1-83, 'A', '85-374, 'RC', <DIX>
A:Cross-references: GB:M44326; NID:G340005; PIDN:AAA61237.1; PID:G553801
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Sun, X.; Skorstengaard, K.; Mosher, D.F.
J. Cell Biol. 118, 693-701, 1992
A:Title: Disulfides modulate RGD-inhibitable cell adhesive activity of thrombospondin
A:Reference number: A42927; MUID:92348511
A:Accession: A42927
A:Molecule type: protein
A:Residues: 987-1003 <SUN>
A:Note: Cys-992 is shown to have a free sulfhydryl
C:Genetics:
A:Gene: GDB:THBS1; TSP1; TSP
A:Map position: 15q15-15q15
A:Cross-references: GDB:120438; OMIM:188060
A:Introns: 23/1
A:Note: the list of introns may be incomplete
C:Complex: homotrimer, disulfide linked
C:Function:
A:Description: participates in cell migration and adhesion, and in platelet aggregati
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;
C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tri

F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-1170/Product: thrombospondin 1 #status predicted <MAT>
 F:317-375/Domain: von Willebrand factor type C repeat homology
 F:378-429/Domain: thrombospondin type 1 repeat homology <THR1>
 F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>
 F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>
 F:551-586/Domain: EGF homology <EGF1>
 F:650-689/Domain: EGF homology <EGF2>
 F:926-928/Region: cell attachment (R-G-D) motif
 F:171-232/Disulfide bonds: #status predicted
 F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:270,274/disulfide bonds: interchain #status predicted
 F:610/Modified site: erythro-beta-hydroxyparagine (Asn) #status predicted
 F:1051/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 94.3%; Score 50; DB 1; Length 1170;
 Best Local Similarity 85.7%; Pred. No. 4.9;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
 |||||
 Db 382 WSPWSEW 388

RESULT 5

A40558
 thrombospondin 1 precursor - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 20-Aug-1999
 C:Accession: A40558; A37905; B42587; S68787
 R:Lawler, J.; Duquette, M.; Ferro, P.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A. Genomics 11, 587-600, 1991
 A:Title: Characterization of the murine thrombospondin gene.
 A:Reference number: A40558; MUID:92128941
 A:Accession: A40558
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1170 <LAW>
 A:Cross-references: GB:M62449; GB:M62450; GB:M62451; GB:M62452; GB:M62453; GB:M62454; GB:M62465; GB:M62466; GB:M62467; GB:M62468; GB:M62469; GB:M62470; NID:G511867; PIDN:AAA5061
 R:Bornstein, P.; Alfi, D.; Devarayalu, S.; Framson, P.; Li, P. J. Biol. Chem. 265, 16691-16698, 1990
 A:Title: Characterization of the mouse thrombospondin gene and evaluation of the role of
 A:Reference number: A37905; MUID:90375546
 A:Accession: A37905
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-490 <BOR>
 A:Cross-references: GB:J05605; GB:J05606; NID:G201991; PIDN:AAA40431.1; PID:G554390
 R:Lahterty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M. J. Biol. Chem. 267, 3274-3281, 1992
 A:Title: Characterization of mouse thrombospondin 2 sequence and expression during cell
 A:Reference number: A42587; MUID:92147683
 A:Accession: B42587
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-1152, 'p', 1154-1170 <LAH>
 A:Cross-references: GB:M87276
 R:Chen, H.; Aeschlimann, D.; Nowlen, J.; Mosher, D.F. FEBS Lett. 387, 36-41, 1996
 A:Title: Expression and initial characterization of recombinant mouse thrombospondin 1 a
 A:Reference number: S68787; MUID:96234006
 A:Accession: S68787
 A:Molecule type: protein
 A:Residues: 19-26, 'x', 28-37 <CHE>
 C:Complex: homotrimer, disulfide linked
 C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; VC
 C:Keywords: calcium binding; glycoprotein; homotrimer
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-1170/Product: thrombospondin 1 #status predicted <MAT>
 F:317-375/Domain: von Willebrand factor type C repeat homology <VWC>

F:378-429/Domain: thrombospondin type 1 repeat homology <THR1>
 F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>
 F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>
 F:551-586/Domain: EGF homology <EGF>
 F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 94.3%; Score 50; DB 2; Length 1170;
 Best Local Similarity 85.7%; Pred. No. 4.9;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
 |||||
 Db 382 WSPWSEW 388

RESULT 6

T25061
 hypothetical protein T21B6.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T25061
 R:Cottage, A. submitted to the EMBL Data Library, November 1995
 A:Reference number: Z19975
 A:Accession: T25061
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-788 <WIL>
 A:Cross-references: EMBL:Z68011; PIDN:CAA92014.1; GSPDB:GN00028; CESP:T21B6.3
 A:Experimental source: clone T21B6
 C:Genetics:
 A:Gene: CESP:T21B6.3
 A:Map position: X
 A:Introns: 20/1; 47/1; 76/1; 152/1; 735/2; 754/2

Query Match 90.6%; Score 48; DB 2; Length 788;
 Best Local Similarity 85.7%; Pred. No. 6.5;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
 |||||
 Db 606 WSPWQEW 612

RESULT 7

A39804
 thrombospondin precursor - chicken
 C:Species: Gallus gallus (chicken)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A39804
 R:Lawler, J.; Duquette, M.; Ferro, P. J. Biol. Chem. 266, 8039-8043, 1991
 A:Title: Cloning and sequencing of chicken thrombospondin.
 A:Reference number: A39804; MUID:91217026
 A:Accession: A39804
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1178 <LAW>
 A:Cross-references: GB:M60853; NID:G212763; PIDN:AAA51437.1; PID:G212764
 C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology
 F:325-383/Domain: von Willebrand factor type C repeat homology <VWC>
 F:386-437/Domain: thrombospondin type 1 repeat homology <THR1>
 F:442-498/Domain: thrombospondin type 1 repeat homology <THR2>
 F:499-555/Domain: thrombospondin type 1 repeat homology <THR3>
 F:658-697/Domain: EGF homology <EGF>

Query Match 88.7%; Score 47; DB 1; Length 1178;
 Best Local Similarity 71.4%; Pred. No. 14;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 WSPWAEW 7
Db 390 WSPWSDW 396

RESULT 8
T09484
cartilage intermediate layer protein precursor - human
C:Species: Homo sapiens (man)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T09484
R:Lorenz, P.; Neame, P.; Sommarin, Y.; Heinegard, D.
J. Biol. Chem. 273, 23469-23475, 1998
A:Title: Cloning and deduced amino acid sequence of a novel cartilage protein (CILP) id
A:Reference number: Z16689; MUID:98389785
A:Accession: T09484
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1184 <LOR>
A:Cross-References: EMBL:AF035408; NID:g3513502; PIDN:AAC33838.1; PID:g3513503
A:Experimental source: tissue type articular cartilage
C:Genetics:
A:Note: CILP
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-1184/Product: cartilage intermediate layer protein #status predicted <MAT>

Query Match 83.0%; Score 44; DB 2; Length 1184;
Best Local Similarity 71.4%; Pred. No. 37;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
Db 152 WSPWSPW 158

RESULT 9
S31002
gene 57 protein - Mycobacterium phage L5
C:Species: Mycobacterium phage L5
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 08-Oct-1999
C:Accession: S31002
R:Donnelly-Wu, W.K.; Jacobs Jr., W.R.; Hatfull, G.F.
Mol. Microbiol. 7, 407-417, 1993
A:Title: Superinfection immunity of mycobacteriophage L5: applications for genetic trans
A:Reference number: S30949; MUID:93211283
A:Accession: S31002
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-152 <DON>
A:Cross-References: EMBL:Z18946; NID:g15859; PIDN:CAA79433.1; PID:g15913
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1992
C:Genetics:
A:Gene: 57

Query Match 81.1%; Score 43; DB 2; Length 152;
Best Local Similarity 71.4%; Pred. No. 7.3;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
Db 13 WSPWRHW 19

RESULT 10
G72806
gp57 protein - Mycobacterium phage D29
C:Species: Mycobacterium phage D29
C>Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 12-Nov-1999
C:Accession: G72806
R:Ford, M.E.; Sarkis, G.J.; Belanger, A.E.; Hendrix, R.W.; Hatfull, G.F.
J. Mol. Biol. 29, 143-164, 1998

```

```

A:Title: Genome structure of mycobacteriophage D29: Implications for phage evolution.
A:Reference number: A72800; MUID:98300335
A:Accession: G72806
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <FOR>
A:Cross-References: GB:AF022214; NID:g3172250; PIDN:AAC18498.1; PID:g3172305
A:Genetics:
A:Gene: 57

Query Match 81.1%; Score 43; DB 2; Length 152;
Best Local Similarity 71.4%; Pred. No. 7.3;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
Db 13 WSPWRHW 19

RESULT 11
T44588
probable transcription regulator tylo [imported] - Streptomyces fradiae
C:Species: Streptomyces fradiae
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 11-May-2000
C:Accession: T44588
R:Bate, N.; Butler, A.R.; Gandecha, A.R.; Cundliffe, E.
Chem. Biol. 6, 617-624, 1999
A:Title: Multiple regulatory genes in the tylosin-biosynthetic cluster of Streptomyce
A:Reference number: Z22801; MUID:99398833
A:Accession: T44588
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-213 <BAT>
A:Cross-References: EMBL:AF145049; PIDN:AAD40803.1
A:Experimental source: strain T59235
C:Genetics:
A:Gene: tylo

Query Match 79.2%; Score 42; DB 2; Length 213;
Best Local Similarity 57.1%; Pred. No. 14;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
Db 120 WAPWRQW 126

RESULT 12
T29247
hypothetical protein F09F9.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29247
R:Minx, P.; Hawkins, J.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid F09F9.
A:Reference number: Z20594
A:Accession: T29247
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-654 <MIN>
A:Cross-References: EMBL:U40958; PIDN:AAA81764.1; CESP:F09F9.4
C:Genetics:
A:Gene: CESP:F09F9.4
A:Introns: 89/3; 138/3; 192/3; 260/1; 312/1; 400/2; 445/3; 496/1; 523/1; 558/2; 596/2

Query Match 79.2%; Score 42; DB 2; Length 654;
Best Local Similarity 71.4%; Pred. No. 41;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 WSPWAEW 7
|||
Db 319 WSAWSEW 325

RESULT 13
T03421
traB protein - Agrobacterium tumefaciens plasmid pTIC58
C:Species: Agrobacterium tumefaciens
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
C:Accession: T03421
R:Piper, K.R.; Beck von Bodman, S.; Cook, D.M.; Hwang, I.; Kim, H.; Farrand, S.K.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z14943
A:Accession: T03421
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-421 <PIP>
A:Cross-references: EMBL:AF010180; NID:g3153171; PIDN:AAC17208.1; PID:g1103912
C:Genetics:
A:Gene: traB
A:Genome: plasmid pTIC58

Query Match 77.4%; Score 41; DB 2; Length 421;
Best Local Similarity 57.1%; Pred. No. 38;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
|||
Db 327 WQPWLOW 333

RESULT 14
B29677
complement C9 precursor - rainbow trout (fragment)
C:Species: Oncorhynchus mykiss (rainbow trout)
C:Date: 31-Dec-1988 #sequence_revision 03-Oct-1995 #text_change 21-Jul-2000
C:Accession: B29677; A56929
R:Stanley, K.K.; Heitz, J.
EMBO J. 6, 1951-1957, 1987
A:Title: Topological mapping of complement component C9 by recombinant DNA techniques su
A:Reference number: A29677; MUID:88004404
A:Accession: B29677
A:Molecule type: mRNA
A:Residues: 1-49, 'V', 51-63, 'A', 65-525, 'RVFLEKGNWSVQPGQAAVEENV' <ST2>
A:Cross-references: EMBL:X05474; NID:g64310; PID:g753800
A:Note: the authors' translation is shown for residues 50 and 64; this sequence has been
R:Tomlinson, S.; Stanley, K.K.; Esser, A.F.
Dev. Comp. Immunol. 17, 67-76, 1993
A:Title: Domain structure, functional activity, and polymerization of trout complement p
A:Reference number: A56929; MUID:93193801
A:Accession: A56929
A:Molecule type: mRNA
A:Residues: 523-574 <TOM>
A:Cross-references: GB:S56372; NID:g266296; PIDN:AAB25728.1; PID:g266297
A:Note: sequence extracted from NCBI backbone (NCBIN:127123, NCBIP:127125)
A:Note: authors' translation is shown for residues 523 and 558
C:Superfamily: complement C9; EGF homology; LDL receptor ligand-binding repeat homology;
C:Keywords: complement alternate pathway; complement pathway; cytolysis; glycoprotein; m
F:17-72/Domain: thrombospondin type 1 repeat homology <THR1>
F:79-112/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:479-510/Domain: EGF homology <EGF>
F:530-574/Domain: thrombospondin type 1 repeat homology <THR2>

Query Match 77.4%; Score 41; DB 2; Length 574;
Best Local Similarity 71.4%; Pred. No. 51;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
|||
Db 21 WSRWSEW 27

RESULT 15
G96518
protein T2E6.8 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G96518
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, L.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: G96518
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-708 <STO>
A:Cross-references: GB:AE005173; NID:g9802586; PIDN:AAF99788.1; GSPDB:GN00141
C:Genetics:
A:Gene: T2E6.8
A:Map position: 1

Query Match 77.4%; Score 41; DB 2; Length 708;
Best Local Similarity 57.1%; Pred. No. 63;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
|||
Db 140 WDPWKDW 146

Search completed: January 31, 2002, 09:24:56
Job time: 261 sec

11

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:19:50 ; Search time 96.28 Seconds
(without alignments)
12.310 Million cell updates/sec

Title: US-09-536-087-6

Perfect score: 84

Sequence: 1 RESHFRGLQNVHLVF 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_ll01:*

- 1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT:*
- 22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84	100.0	16	AA19679	Human thrombospondin
2	84	100.0	151	AA19679	Human colon cancer
3	84	100.0	1172	AA19677	Human thrombospondin
4	84	100.0	1172	AA19677	Human thrombospondin
5	51	60.7	12	AA19677	Human thrombospondin
6	44	52.4	459	AA19677	Alpha3beta1 integrin
7	44	52.4	459	AA19677	Angiotensin convert
8	44	52.4	459	AA19677	Human cancer assoc
9	44	52.4	555	AA19677	Angiotensin convert
10	44	52.4	616	AA19677	Angiotensin convert
11	44	52.4	731	AA19677	Human protein sequ
					Angiotensin convert

12	44	52.4	1152	21	AA19677	Human thrombospondin
13	44	52.4	1170	22	AA19677	Human shear stress
14	44	52.4	1170	22	AA19677	Human variant thro
15	43	51.2	12	22	AA19677	Alpha3beta1 integrin
16	43	51.2	12	22	AA19677	Alpha3beta1 integrin
17	43	51.2	12	22	AA19677	Alpha3beta1 integrin
18	42	50.0	12	22	AA19677	Alpha3beta1 integrin
19	42	50.0	12	22	AA19677	Alpha3beta1 integrin
20	42	50.0	468	20	AA19677	Human PRO865 prote
21	42	50.0	468	21	AA19677	Human PRO865 prote
22	42	50.0	468	21	AA19677	Human secreted pro
23	42	50.0	468	21	AA19677	Human PRO865 prote
24	42	50.0	468	21	AA19677	Human signal pepti
25	42	50.0	468	22	AA19677	Secreted protein e
26	42	50.0	468	22	AA19677	Human secreted pro
27	42	50.0	469	20	AA19677	Mammalian Erol pro
28	42	50.0	481	21	AA19677	Human Th2-specific
29	42	50.0	494	22	AA19677	Gene #6 associated
30	42	50.0	497	22	AA19677	Gene #28 associate
31	42	50.0	552	22	AA19677	Human colon cancer
32	41	48.8	12	22	AA19677	Alpha3beta1 integrin
33	41	48.8	12	22	AA19677	Alpha3beta1 integrin
34	41	48.8	464	20	AA19677	Hypoxia-regulated
35	41	48.8	464	21	AA19677	Polypeptide isolat
36	41	48.8	464	21	AA19677	Murine Th2-specifi
37	41	48.8	464	22	AA19677	Mouse endoplasmic
38	40	47.6	12	22	AA19677	Alpha3beta1 integrin
39	39	46.4	12	22	AA19677	Alpha3beta1 integrin
40	39	46.4	126	22	AA19677	Carboxyphosphoenol
41	39	46.4	364	14	AA19677	Mutant PDG encoded
42	39	46.4	365	14	AA19677	Mutant PDG encoded
43	39	46.4	374	14	AA19677	Mutant PDG encoded
44	39	46.4	395	14	AA19677	Mutant PDG encoded
45	39	46.4	403	14	AA19677	Mutant PDG encoded

ALIGNMENTS

RESULT 1

AA19679
ID AA19679 standard; Peptide; 16 AA.

AC AA19679;

XX 05-FEB-2001 (first entry)

XX Human thrombospondin-2 peptide 1 (aa198-213).

DE Thrombospondin-2; TSP-2; human; angiogenesis; cell proliferation;
KW melanoma; tumour; cancer; squamous cell carcinoma; antiangiogenic;
KW prostate cancer; psoriasis; rosecea dermatosis; antitumour;
KW therapy.

OS Homo sapiens.

XX WO200057899-A1.

PN 05-OCT-2000.

PD 24-MAR-2000; 2000WO-US07835.

XX 31-MAR-1999; 99US-0127221.

XX (GEO) GEN HOSPITAL CORP.

PI Detmar M, Streit M;

XX WPI; 2000-656131/63.

XX Treating a disorder characterized by unwanted cell proliferation e.g.
PT precancerous, cancerous or neoplastic cells or presence of tumor
PT preferably of skin or prostate, comprises increasing thrombospondin-2

PT activity -
 PS Disclosure; Page 40; 73pp; English.
 XX
 CC The present sequence is that of peptide 1, derived from the
 CC procollagen domain (amino acids 198-213) of human thrombospondin-2
 CC (TSP-2, see AAB19677). The peptide is 1 of 5 synthetic peptides (see
 CC AAB19679-83) used with human dermal microvascular endothelial cells
 CC (HMEC) to determine the effect of TSP-2 on endothelial cell
 CC migration. Peptide 1 did not significantly modify HMEC migration,
 CC in contrast to peptide 7 (see AAB19683) from a type 1 repeat of
 CC TSP-2, which inhibited migration. The invention is based on the
 CC discovery that overexpression of TSP-2 decreases tumor size in
 CC vivo, and features methods for modulating unwanted angiogenesis and
 CC cell proliferation by increasing TSP-2 activity.
 XX
 SQ Sequence 16 AA;
 Query Match 100.0%; Score 84; DB 21; Length 16;
 Best Local Similarity 100.0%; Pred. No. 9.8e-08;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RESHFRGLLNQVHLVF 16
 Db |||||
 1 reshfrgllqnvhlf 16
 |||||
 RESULT 2
 AAB53711
 ID AAB53711 standard; Protein; 151 AA.
 AC AAB53711;
 XX
 DT 09-MAR-2001 (first entry)
 DE Human colon cancer antigen protein sequence SEQ ID NO:1251.
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioactive; neuroprotective; vulnery;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder.
 OS Homo sapiens.
 XX
 PN WO200055351-A1.
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05883.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-587534/55.
 DR N-PSDB; AAC98468.
 XX
 PT Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -
 XX
 PS Claim 11; Page 1837-1838; 2104pp; English.
 XX
 CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,

CC vulnery, nephrotropic, antiinfective and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.
 XX
 SQ Sequence 151 AA;
 Query Match 100.0%; Score 84; DB 21; Length 151;
 Best Local Similarity 100.0%; Pred. No. 1.1e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RESHFRGLLNQVHLVF 16
 Db |||||
 84 reshfrgllqnvhlf 99
 |||||
 RESULT 3
 AAB19677
 ID AAB19677 standard; Protein; 1172 AA.
 AC AAB19677;
 XX
 DT 05-FEB-2001 (first entry)
 DE Human thrombospondin-2.
 XX
 KW Thrombospondin-2; TSP-2; human; angiogenesis; cell proliferation;
 KW melanoma; tumour; cancer; squamous cell carcinoma; antiangiogenic;
 KW prostate cancer; psoriasis; rosecea dermatosis; antitumour;
 KW therapy.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 382..429
 FT /note= "type 1 repeat"
 FT Region 384..390
 FT /note= "WSPWAEW sequence involved in antiangiogenic
 FT activity"
 FT Region 438..490
 FT /note= "type 1 repeat"
 FT Region 495..547
 FT /note= "type 1 repeat"
 XX
 PN WO200057899-A1.
 XX
 PD 05-OCT-2000.
 XX
 PF 24-MAR-2000; 2000WO-US07835.
 XX
 PR 31-MAR-1999; 99US-0127221.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Detmar M, Streit M;
 XX
 DR WPI; 2000-656131/63.
 DR N-PSDB; AAA88669.
 XX
 PT Treating a disorder characterized by unwanted cell proliferation e.g.
 PT precancerous, cancerous or neoplastic cells or presence of tumor
 PT preferably of skin or prostate, comprises increasing thrombospondin-2
 PT activity -
 XX

PS Disclosure; Fig 2; 73pp; English.

XX The present sequence is that of human thrombospondin-2 (TSP-2).
 CC The invention is based on the discovery that overexpression of
 CC TSP-2 decreases tumor size in vivo, and features methods for
 CC modulating unwanted angiogenesis and tumour growth. Treatment of
 CC unwanted cell proliferation or angiogenesis involves increasing
 CC TSP-2 activity. This is achieved by administering an agent which
 CC increases TSP-2 activity, especially a TSP-2 polypeptide, a TSP-2
 CC derived polypeptide or retro-inverso peptide, a nucleic acid
 CC encoding TSP-2, an agonist of TSP-2, or an agent that increases
 CC TSP-2 gene expression. The TSP-2 polypeptide may include at least
 CC one type 1 repeat such as the WSP/AEW peptide (see AAB19683),
 CC which is involved in the antiangiogenic activity of TSP-2. The
 CC method is used to treat a disorder characterised by pre-cancerous,
 CC cancerous or neoplastic cells, or the presence of a tumour, or a
 CC disorder that affects epithelial tissues resulting in unwanted
 CC skin cell proliferation. Such disorders include malignant
 CC melanoma, prostate cancer, squamous cell carcinoma, aged skin,
 CC rosacea dermatosis, psoriasis, and skin damage caused by
 CC photoradiation (all claimed). Evaluating the presence of TSP-2
 CC nucleic acid or protein is useful for diagnosing a subject at risk
 CC of unwanted cell proliferation or angiogenesis. Methods are also
 CC provided of identifying compounds that modulate TSP-2 activity.

XX SQ Sequence 1172 AA;

Query Match 100.0%; Score 84; DB 21; Length 1172;
 Best Local Similarity 100.0%; Pred. No. 9.8e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESHFRGLQNVLHF 16
 |||||
 Db 198 reshfrgllqnvhlf 213

RESULT 4
 AAB00043
 ID AAB00043 standard; Protein; 1172 AA.

XX AC AAB00043;
 XX DT 08-NOV-2000 (first entry)
 XX DE Human thrombospondon-2 (TSP-2).
 XX KW TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
 KW thrombospondin; angiogenesis; tumour; treatment; cancer;
 KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
 KW glaucoma.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT Region 381..436
 FT /label= Type 1 repeat region
 FT Region 437..493
 FT /label= Type 1 repeat region
 FT Region 494..550
 FT /label= Type 1 repeat region
 XX PN W0200044908-A2.
 XX PD 03-AUG-2000.
 XX PF 01-FEB-2000; 2000WO-US02482.
 XX PR 01-FEB-1999; 99US-0118053.
 XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 XX PI Lawler JW;

XX WPI; 2000-514823/46.

XX Nucleic acids encoding chimeric proteins such as cartilage oligomeric
 PT matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for
 PT inhibiting angiogenesis and treating diseases such as cancer
 XX Disclosure; Fig 2; 40pp; English.

XX New nucleic acids are described which encode a protein comprising
 CC the second and third type 1 repeats of human TSP (thrombospondin)-1,
 CC but not the TGF (transforming growth factor)-beta activation region
 CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing
 CC the second and third type-1 repeats and the COMP (cartilage
 CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was
 CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1
 CC caused inhibition of the growth of tumours in mice models.
 CC Thus the nucleic acids and proteins may be useful for treating
 CC angiogenesis related diseases such as cancer (by reducing the rate of
 CC growth and size of tumours), arthritis, psoriasis, diabetic
 CC retinopathy, corneal graft rejection, and glaucoma. They may also be
 CC used for treating human immunodeficiency virus (HIV) infection.
 CC Anti-angiogenic therapy has little toxicity, does not require the
 CC therapeutic agent to enter tumour cells or cross the blood-brain
 CC barrier, controls tumour growth independently of growth of
 CC tumour cell heterogeneity, and does not induce drug resistance.

XX SQ Sequence 1172 AA;

Query Match 100.0%; Score 84; DB 21; Length 1172;
 Best Local Similarity 100.0%; Pred. No. 9.8e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESHFRGLQNVLHF 16
 |||||
 Db 198 reshfrgllqnvhlf 213

RESULT 5
 AAB35367
 ID AAB35367 standard; Peptide; 12 AA.

XX AC AAB35367;
 XX DT 08-MAY-2001 (first entry)
 XX DE Alpha3beta1 integrin binding peptide #32.
 XX KW Alpha3beta1 integrin; angiogenesis; cell proliferation; cancer;
 KW diabetic retinopathy; restenosis; atherosclerosis; rheumatoid arthritis;
 KW macular degeneration; psoriasis; cell adhesion; cell motility.
 XX OS Synthetic.
 XX PN W0200105812-A2.
 XX PD 25-JAN-2001.
 XX PF 12-JUL-2000; 2000WO-US18986.
 XX PR 15-JUL-1999; 99US-0144549.
 XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX PI Roberts DD, Krutzsch HC;
 XX WPI; 2001-182656/18.
 XX New peptides that bind to or are recognized by alpha3-beta1 integrins,
 PT useful for inhibiting cell adhesion to extracellular matrix, cell
 PT motility and proliferation and for treating rheumatoid arthritis and
 PT cancer

XX Example 2; Page 34; 84pp; English.

XX The present invention provides a number of peptides which bind to

CC alpha2beta1 integrins. They are useful in the modulation of cell adhesion

CC and motility, and in the treatment of cancer, diabetic retinopathy,

CC rheumatoid arthritis, macular degeneration, atherosclerosis, psoriasis

CC and restenosis. The present sequence is an example of one of the peptides

CC of the invention.

XX Sequence 12 AA;

SQ

Query Match 60.7%; Score 51; DB 22; Length 12;

Best Local Similarity 75.0%; Pred. No. 0.026;

Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 5 PRGLLQNVHLVF 16

Db 1 fggvlnvhnfvf 12

RESULT 6

AAU02916

ID AAU02916 standard; Protein; 459 AA.

XX

AC AAU02916;

DT 12-SEP-2001 (first entry)

XX

DE Angiotensin converting enzyme (ACEV) splice variant protein #16.

XX

KW Angiotensin converting enzyme splice variant; ACEV; interleukin 6;

KW granulocyte colony stimulating factor receptor; glucagon; hypertrophy;

KW platelet-derived endothelial cell growth factor; cardiovascular disease;

KW cellular tumour antigen P53; cyclin-dependent kinase inhibitor 1C;

KW vasoactive intestinal polypeptide receptor 2; arteriosclerosis; cancer;

KW myocardial infarction; coronary arterial thrombosis; renal disease;

KW diabetic nephropathy; muscular disease; immune disorder; sarcoidosis;

KW multiple sclerosis; immune complex nephritis; deep vein thrombosis;

KW nonaroidotic pulmonary granulomatous disease; endothelial abnormality;

KW vascular disorder; asbestosis.

XX

OS Homo sapiens.

XX

PN WO200136632-A2.

XX

PD 25-MAY-2001.

XX

PF 17-NOV-2000; 2000WO-IL00766.

XX

PR 17-NOV-1999; 99IL-0132978.

PR 10-DEC-1999; 99IL-0133455.

XX

PA (COMP-) COMPUGEN LTD.

XX

XX Levine Z, David A, Azar I, Khosravi R, Bernstein J;

PI

XX WPI; 2001-336004/35.

DR

DR N-PSDB; AAS06016.

XX

XX Novel alternative splicing variants e.g. variant of angiotensin

PT converting enzyme (ACEV), useful in identifying candidate compounds

PT capable of binding to the variant and to detect anti-variant antibodies

PT

XX

PS Claim 4; Fig 16; 519pp; English.

XX

XX The sequence represents an angiotensin converting enzyme splice variant

CC (ACEV) polypeptide. The polypeptides of the invention include variants of

CC granulocyte colony stimulating factor receptor, glucagon, interleukin 6,

CC platelet-derived endothelial cell growth factor, cyclin-dependent kinase

CC inhibitor 1C, cellular tumour antigen P53, and vasoactive intestinal

CC polypeptide receptor 2. The polypeptides and their associated nucleic

CC acids are useful for identification of variant sequences and detection of

CC candidate compounds capable of binding the molecules. The sequences of

CC the invention can be used in the treatment and diagnosis of various

CC disorders including cardiovascular diseases such as arteriosclerosis,

CC myocardial infarction and coronary arterial thrombosis, renal diseases

CC such as diabetic nephropathy, muscular diseases such as hypertrophy,

CC immune disorders such as immune complex nephritis, multiple sclerosis,

CC cancer, sarcoidosis, nonaroidotic pulmonary granulomatous diseases such

CC as asbestosis and vascular pathologies involving an endothelial

CC abnormality such as deep vein thrombosis.

XX Sequence 459 AA;

SQ

Query Match 52.4%; Score 44; DB 22; Length 459;

Best Local Similarity 61.5%; Pred. No. 20;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 4 HFRGLLQNVHLVF 16

Db 207 nfgvlnvhnfvf 219

RESULT 7

AAU03602

ID AAB43602 standard; Protein; 466 AA.

XX

AC AAB43602;

DT 08-FEB-2001 (first entry)

XX

DE Human cancer associated protein sequence SEQ ID NO:1047.

XX

KW Human; cancer associated gene; cancer antigen; detection; cancer;

KW diagnosis; cytostatic; proliferative; vulnery; immunomodulator;

KW antidiabetic; antisthmatic; antirheumatic; antiarthritic; antiviral;

KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;

KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;

KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;

KW immune disorder; haematopoietic cell disorder; autoimmune disorder;

KW allergic reaction; graft versus host disease; organ rejection;

KW haemostatic; thrombolytic; cardiovascular disorder; infection;

KW neurological disease; drug screening.

XX

OS Homo sapiens.

XX

PN WO200055350-A1.

XX

PD 21-SEP-2000.

XX

PF 08-MAR-2000; 2000WO-US05882.

XX

PR 12-MAR-1999; 99US-0124270.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

XX Rosen CA, Ruben SM;

PI

XX WPI; 2000-587533/55.

DR

DR N-PSDB; AAC77811.

XX

XX Novel isolated nucleic acids comprising sequences encoding peptides

PT useful for treating or diagnosing e.g. cancer

PT

XX

PS Claim 11; Page 1636-1638; 2352pp; English.

XX

XX AAC77607 to AAC78448 encode the human cancer associated proteins given

CC in AAB43398 to AAB44239. The proteins can have activities based on the

CC tissues and cells the genes are expressed in. Example of activities

CC include: cytostatic; proliferative; vulnery; immunomodulator;

CC antidiabetic; antisthmatic; antirheumatic; antiarthritic;

CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;

CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
 CC nontropic; vasotropic; antipsoriatic and antiangiogenic. The
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention.

XX Sequence 466 AA;

Query Match 52.4%; Score 44; DB 21; Length 466;
 Best Local Similarity 61.5%; Pred. No. 20;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 HFRGLLQNVHLVF 16
 Db 260 nfggvlqnvrvfvf 272
 :|:|:|:|:|:|:|

RESULT 8

AAU02915
 ID AAU02915 standard; Protein; 546 AA.

XX AAU02915;

XX 12-SEP-2001 (first entry)

XX Angiotensin converting enzyme (ACEV) splice variant protein #15.

XX Angiotensin converting enzyme splice variant; ACEV; interleukin 6;
 KW granulocyte colony stimulating factor receptor; glucagon; hypertrophy;
 KW platelet-derived endothelial cell growth factor; cardiovascular disease;
 KW cellular tumour antigen P53; cyclin-dependent kinase inhibitor 1C;
 KW vasoactive intestinal polypeptide receptor 2; arteriosclerosis; cancer;
 KW myocardial infarction; coronary arterial thrombosis; renal disease;
 KW diabetic nephropathy; muscular disease; immune disorder; sarcoidosis;
 KW multiple sclerosis; immune complex nephritis; deep vein thrombosis;
 KW nonaroidotic pulmonary granulomatous disease; endothelial abnormality;
 KW vascular disorder; asbestosis.

XX Homo sapiens.

XX WO200136632-A2.

XX 25-MAY-2001.

XX 17-NOV-2000; 2000WO-IL00766.

XX 17-NOV-1999; 99IL-0132978.

XX 10-DEC-1999; 99IL-0133455.

XX (COMP-) COMPUGEN LTD.

XX Levine Z, David A, Azar I, Khosravi R, Bernstein J;

XX WPI; 2001-336004/35.

XX N-PSDB; AAS06015.

XX Novel alternative splicing variants e.g. variant of angiotensin
 PT converting enzyme (ACEV), useful in identifying candidate compounds
 PT capable of binding to the variant and to detect anti-variant antibodies

XX Claim 4; Fig 15; 519pp; English.

XX The sequence represents an angiotensin converting enzyme splice variant
 CC (ACEV) polypeptide. The polypeptides of the invention include variants of
 CC granulocyte colony stimulating factor receptor, glucagon, interleukin 6,
 CC platelet-derived endothelial cell growth factor, cyclin-dependent kinase
 CC inhibitor 1C, cellular tumour antigen P53, and vasoactive intestinal
 CC polypeptide receptor 2. The polypeptides and their associated nucleic
 CC acids are useful for identification of variant sequences and detection of
 CC candidate compounds capable of binding the molecules. The sequences of
 CC the invention can be used in the treatment and diagnosis of various
 CC disorders including cardiovascular diseases such as arteriosclerosis,
 CC myocardial infarction and coronary arterial thrombosis, renal diseases
 CC such as diabetic nephropathy, muscular diseases such as hypertrophy,
 CC immune disorders such as immune complex nephritis, multiple sclerosis,
 CC cancer, sarcoidosis, nonaroidotic pulmonary granulomatous diseases such
 CC as asbestosis and vascular pathologies involving an endothelial
 CC abnormality such as deep vein thrombosis.

XX Sequence 546 AA;

Query Match 52.4%; Score 44; DB 22; Length 546;
 Best Local Similarity 61.5%; Pred. No. 24;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 HFRGLLQNVHLVF 16
 Db 207 nfggvlqnvrvfvf 219
 :|:|:|:|:|:|:|

RESULT 9

AAU02914
 ID AAU02914 standard; Protein; 555 AA.

XX AAU02914;

XX 12-SEP-2001 (first entry)

XX Angiotensin converting enzyme (ACEV) splice variant protein #14.

XX Angiotensin converting enzyme splice variant; ACEV; interleukin 6;
 KW granulocyte colony stimulating factor receptor; glucagon; hypertrophy;
 KW platelet-derived endothelial cell growth factor; cardiovascular disease;
 KW cellular tumour antigen P53; cyclin-dependent kinase inhibitor 1C;
 KW vasoactive intestinal polypeptide receptor 2; arteriosclerosis; cancer;
 KW myocardial infarction; coronary arterial thrombosis; renal disease;
 KW diabetic nephropathy; muscular disease; immune disorder; sarcoidosis;
 KW multiple sclerosis; immune complex nephritis; deep vein thrombosis;
 KW nonaroidotic pulmonary granulomatous disease; endothelial abnormality;
 KW vascular disorder; asbestosis.

XX Homo sapiens.

XX WO200136632-A2.

XX 25-MAY-2001.

XX 17-NOV-2000; 2000WO-IL00766.

XX 17-NOV-1999; 99IL-0132978.

XX 10-DEC-1999; 99IL-0133455.

XX (COMP-) COMPUGEN LTD.

XX Levine Z, David A, Azar I, Khosravi R, Bernstein J;

XX WPI; 2001-336004/35.

XX N-PSDB; AAS06014.

XX Novel alternative splicing variants e.g. variant of angiotensin
 PT converting enzyme (ACEV), useful in identifying candidate compounds
 PT capable of binding to the variant and to detect anti-variant antibodies

XX PS Claim 4; Fig 14; 519pp; English.

XX CC The sequence represents an angiotensin converting enzyme splice variant

XX CC (ACEV) polypeptide. The polypeptides of the invention include variants of

XX CC granulocyte colony stimulating factor receptor, glucagon, interleukin 6,

XX CC platelet-derived endothelial cell growth factor, cyclin-dependent kinase

XX CC inhibitor 1C, cellular tumour antigen p53, and vasoactive intestinal

XX CC polypeptide receptor 2. The polypeptides and their associated nucleic

XX CC acids are useful for identification of variant sequences and detection of

XX CC candidate compounds capable of binding the molecules. The sequences of

XX CC the invention can be used in the treatment and diagnosis of various

XX CC disorders including cardiovascular diseases such as arteriosclerosis,

XX CC myocardial infarction and coronary arterial thrombosis, renal diseases

XX CC such as diabetic nephropathy, muscular diseases such as hypertrophy,

XX CC immune disorders such as immune complex nephritis, multiple sclerosis,

XX CC cancer, sarcoidosis, nonaroidotic pulmonary granulomatous diseases such

XX CC as asbestosis and vascular pathologies involving an endothelial

XX CC abnormality such as deep vein thrombosis.

XX SQ Sequence 555 AA;

Query Match 52.4%; Score 44; DB 22; Length 555;

Best Local Similarity 61.5%; Pred. No. 24;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 HFRGLQNVHLVF 16

Db 207 nfgvgvqnrvivf 219

RESULT 10

AA095123

ID AAB95123 standard; Protein: 616 AA.

XX AC AAB95123;

XX DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:17120.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX DR WPI; 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602

XX PT full-length cDNAs defined in the specification, and for the detection

XX PT and/or diagnosis of the abnormality of the proteins encoded by the

XX PT full-length cDNAs -

XX PS Claim 8; SEQ ID 17120; 2537pp + CD ROM; English.

XX CC The present invention describes primer sets for synthesising 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesising polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX SQ Sequence 616 AA;

Query Match 52.4%; Score 44; DB 22; Length 616;

Best Local Similarity 53.8%; Pred. No. 27;

Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RESHFRGLQNVH 13

Db 224 reaharqvlenih 236

RESULT 11

AA02913

ID AAU02913 standard; Protein: 731 AA.

XX AC AAU02913;

XX DT 12-SEP-2001 (first entry)

XX DE Angiotensin converting enzyme (ACEV) splice variant protein #13.

XX KW Angiotensin converting enzyme splice variant; ACEV; interleukin 6;

XX KW granulocyte colony stimulating factor receptor; glucagon; hypertrophy;

XX KW platelet-derived endothelial cell growth factor; cardiovascular disease;

XX KW cellular tumour antigen p53; cyclin-dependent kinase inhibitor 1C;

XX KW vasoactive intestinal polypeptide receptor 2; arteriosclerosis; cancer;

XX KW myocardial infarction; coronary arterial thrombosis; renal disease;

XX KW diabetic nephropathy; muscular disease; immune disorder; sarcoidosis;

XX KW multiple sclerosis; immune complex nephritis; deep vein thrombosis;

XX KW nonaroidotic pulmonary granulomatous disease; endothelial abnormality;

XX KW vascular disorder; asbestosis.

XX OS Homo sapiens.

XX PN WO200136632-A2.

XX PD 25-MAY-2001.

XX PI 17-NOV-2000; 2000WO-IL00766.

XX PR 17-NOV-1999; 99IL-0132978.

XX PR 10-DEC-1999; 99IL-0133455.

XX PA (COMP-) COMPUGEN LTD.

XX PI Levine Z, David A, Azar I, Khosravi R, Bernstein J;

XX DR WPI; 2001-336004/35.

XX DR N-PSDB; AAS06013.

XX PS Claim 60; Page 515-521; 678pp; Japanese.

XX CC The present invention provides the protein and coding sequences of a

CC number of human shear stress response proteins. These are useful in the

CC diagnosis, treatment and screening of vascular diseases caused by

CC arteriosclerosis, including heart failure, post-PTCA restenosis and

CC hypertension.

XX SQ Sequence 1170 AA;

Query Match 52.4%; Score 44; DB 22; Length 1170;

Best Local Similarity 61.5%; Pred. No. 54;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 HFRGLQNVLVLF 16

Db 207 nfgvlnvrvfvf 219

RESULT 14

AAB74450

ID AAB74450 standard; Protein; 1170 AA.

AC AAB74450;

DT 06-JUN-2001 (first entry)

DE Human variant thrombospondin 1.

XX KW Human; variant thrombospondin 1; variant thrombospondin 4; SNP;

KW polymorphism; vascular disease; coronary artery disease; forensics;

KW myocardial infarction; atherosclerosis; stroke; venous thromboembolism;

KW pulmonary embolism; paternity test.

XX OS Homo sapiens.

PN WO200118250-A2.

PD 15-MAR-2001.

XX PF 07-SEP-2000; 2000WO-US24503.

XX PR 10-SEP-1999; 99US-0153357.

PR 26-JUL-2000; 2000US-0220947.

PR 16-AUG-2000; 2000US-0225724.

XX PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

PA (MILL-) MILLENNIUM PHARM INC.

XX PI Lander ES, Gargill M, Ireland JS, Bolk S, Daley GQ, McCarthy JJ;

XX WPI; 2001-226749/23.

DR N-PSDB; AAF95238.

XX PT Nucleic acids comprising single nucleotide polymorphisms, useful in

PT applications such as forensics, paternity testing, medicine, genetic

PT analysis and phenotype correlations to diseases such as diabetes and

PT atherosclerosis.

XX PS Claim 20; Fig 1; 242pp; English.

XX CC The present invention provides a method of diagnosing a vascular disease

CC in an individual, involving determining the sequence at various

CC polymorphic sites within the human thrombospondin 1 and thrombospondin 4

CC genes. The sequences at a number of polymorphic sites are also provided

CC in the specification. In particular, the method can be used in the

CC diagnosis of atherosclerosis, myocardial infarction, coronary heart

CC disease, stroke, peripheral vascular diseases, venous thromboembolism

CC and pulmonary embolism. Single nucleotide polymorphisms (SNPs) are also

CC useful in forensics, paternity testing, genetic analysis and phenotype

CC correlations to diseases. The present sequence is the human variant

CC thrombospondin 1 protein.

XX SQ Sequence 1170 AA;

Query Match 52.4%; Score 44; DB 22; Length 1170;

Best Local Similarity 61.5%; Pred. No. 54;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 HFRGLQNVLVLF 16

Db 207 nfgvlnvrvfvf 219

RESULT 15

AAB35352

ID AAB35352 standard; Peptide; 12 AA.

AC AAB35352;

DT 08-MAY-2001 (first entry)

DE Alpha3beta1 integrin binding peptide #17.

XX KW Alpha3beta1 integrin; angiogenesis; cell proliferation; cancer;

KW diabetic retinopathy; restenosis; atherosclerosis; rheumatoid arthritis;

KW macular degeneration; psoriasis; cell adhesion; cell motility.

XX OS Synthetic.

PN WO200105812-A2.

XX PD 25-JAN-2001.

XX PF 12-JUL-2000; 2000WO-US18986.

XX PR 15-JUL-1999; 99US-0144549.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PI Roberts DD, Krutzsch HC;

XX DR WPI; 2001-182656/18.

XX PT New peptides that bind to or are recognized by alpha3-beta1 integrins,

PT useful for inhibiting cell adhesion to extracellular matrix, cell

PT motility and proliferation and for treating rheumatoid arthritis and

PT cancer.

XX PS Claim 4; Page 34; 84pp; English.

XX CC The present invention provides a number of peptides which bind to

CC alpha3beta1 integrins. They are useful in the modulation of cell adhesion

CC and motility, and in the treatment of cancer, diabetic retinopathy,

CC rheumatoid arthritis, macular degeneration, atherosclerosis, psoriasis

CC and restenosis. The present sequence is an example of one of the peptides

CC of the invention.

XX SQ Sequence 12 AA;

Query Match 51.2%; Score 43; DB 22; Length 12;

Best Local Similarity 66.7%; Pred. No. 0.58;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 FRGLQNVLVLF 16

Db 1 fggvlnvrvfvf 12

Search completed: January 31, 2002, 09:23:04

Job time: 194 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:24:01 ; Search time 28.31 seconds
(without alignments)
20.722 Million cell updates/sec

Title: US-09-536-087-6

Perfect score: 84

Sequence: 1 RESHFRGLQNVLHVF 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	84	100.0	1172	1	TS22_HUMAN
2	84	100.0	1172	1	TS22_MOUSE
3	79	94.0	1178	1	TS22_CHICK
4	78	92.9	1170	1	TS22_BOVIN
5	45	53.6	1173	1	TS21_XENLA
6	44	52.4	1170	1	TS21_BOVIN
7	44	52.4	1170	1	TS21_HUMAN
8	44	52.4	1170	1	TS21_MOUSE
9	42	50.0	361	1	YKYL_CAEEL
10	40	47.6	397	1	YR23_MYCTU
11	40	47.6	501	1	C723_ARATH
12	40	47.6	816	1	NER2_ECOLI
13	39	46.4	409	1	SER2_MOUSE
14	38.5	45.8	4466	1	DYHC_ANTCR
15	38.5	45.8	4466	1	DYHC_TRIGR
16	38	45.2	351	1	FTSZ_THEMEA
17	38	45.2	525	1	YMD6_RHIME
18	38	45.2	622	1	FEF5_YEAST
19	38	45.2	750	1	DPOL_HPBVZ
20	38	45.2	856	1	TTK_MOUSE
21	38	45.2	1417	1	BLM_HUMAN
22	37.5	44.6	641	1	YGG0_YEAST
23	37	44.0	173	1	ATP7_KLULA
24	37	44.0	212	1	PUR3_ECOLI
25	37	44.0	252	1	TH14_PIRAB
26	37	44.0	255	1	TH14_PYRHO
27	37	44.0	340	1	LPSE_RHIME
28	37	44.0	482	1	CATA_VIBFI
29	37	44.0	630	1	PARE_SALTY
30	37	44.0	650	1	Y096_MYCGE
31	37	44.0	651	1	BGLR_CANFA
32	37	44.0	651	1	BGLR_FELCA
33	37	44.0	909	1	CSRP_RAT

ALIGNMENTS

RESULT 1

ID	TSP2_HUMAN	STANDARD;	PRT;	1172 AA.
AC	P35442;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	01-AUG-2001 (Rel. 40, Last annotation update)			
DE	THROMBOSPONDIN 2 PRECURSOR.			
GN	THBS2 OR TSP2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94010892; PubMed=8406456;			
RA	Labell T.L., Byers P.H.;			
RT	"Sequence and characterization of the complete human thrombospondin 2			
RT	cDNA: potential regulatory role for the 3' untranslated region.";			
RL	Genomics 17:225-229(1993).			
RN	[2]			
RP	SEQUENCE OF 560-1172 FROM N.A.			
RC	TISSUE=Fibroblast;			
RX	MEDLINE=92217961; PubMed=1559694;			
RA	Labell T.L., McGookey Milewicz D.J., Distche C.M., Byers P.H.;			
RT	"Thrombospondin II: partial cDNA sequence, chromosome location, and			
RT	expression of a second member of the thrombospondin gene family in			
RT	humans.";			
RL	Genomics 12:421-429(1992).			
CC	-1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND			
CC	CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,			
CC	LAMININ AND TYPE V COLLAGEN.			
CC	-1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.			
CC	-1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.			
CC	-1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.			
CC	-1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; L12350; AAA03703.1; -			
DR	EMBL; M81339; -; NOT_ANNOTATED_CDS.			
DR	PIR; A42173; A42173.			
DR	HSSP; P00740; 11XA.			
DR	MIM; 188061; -			
DR	InterPro; IPR000561; EGF-like.			
DR	InterPro; IPR001881; EGF-Ca.			
DR	InterPro; IPR001791; Laminin_G.			
DR	InterPro; IPR000884; TSPl.			

059610 pyrococcus
Q57992 methanococc
O83103 treponema p
P56061 helicobacte
P38074 saccharomyc
O28390 archaeoglob
P77808 escherichia
P15273 versinia en
P08538 versinia ps
P70718 actinobacil
P43774 haemophilus
P24461 oryctolagus

34 37 44.0 1235 1 DPOL_PYRHO
35 36 42.9 86 1 Y572_METJA
36 36 42.9 191 1 Y064_TREPA
37 36 42.9 276 1 PANC_HELPY
38 36 42.9 348 1 HMT1_YEAST
39 36 42.9 380 1 RPA2_ARCFU
40 36 42.9 400 1 CINA_ECOLI
41 36 42.9 468 1 YOPH_YEREN
42 36 42.9 468 1 YOPH_YERPS
43 36 42.9 484 1 6PGD_ACTAC
44 36 42.9 484 1 6PGD_HAEIN
45 36 42.9 494 1 CPGL_RABIT

DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; WMFC.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 9.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00092; TSP1; 3.
DR PROSITE; PS01208; WMFC; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1172
FT DOMAIN 19 232
FT DOMAIN 318 375
FT DOMAIN 381 432
FT DOMAIN 437 493
FT DOMAIN 494 548
FT DOMAIN 549 589
FT DOMAIN 590 647
FT DOMAIN 648 692
FT DOMAIN 725 760
FT DOMAIN 761 783
FT DOMAIN 784 819
FT DOMAIN 820 842
FT DOMAIN 843 880
FT DOMAIN 881 916
FT DOMAIN 917 952
FT DOMAIN 953 1172
FT SITE 928 930
FT DISULFID 266 266
FT DISULFID 270 270
FT DISULFID 553 564
FT DISULFID 558 574
FT DISULFID 577 588
FT DISULFID 594 610
FT DISULFID 601 619
FT DISULFID 622 646
FT DISULFID 652 665
FT DISULFID 659 678
FT DISULFID 680 691
FT CARBOHYD 151 151
FT CARBOHYD 316 316
FT CARBOHYD 330 330
FT CARBOHYD 457 457
FT CARBOHYD 584 584
FT CARBOHYD 710 710
FT CARBOHYD 1069 1069
SQ SEQUENCE 1172 AA; 129955 MW; 2AC7BB230E44C6F5 CRC64;
Query Match 100.0%; Score 84; DB 1; Length 1172;
Best Local Similarity 100.0%; Pred. No. 1.le-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 RESHFRGLQNHLVLF 16
|||||
Db 198 RESHFRGLQNHLVLF 213
RESULT 2
TSP2_MOUSE STANDARD; PRT; 1172 AA.
AC 003350;
DT 01-JUN-1994 (Rel. 29, Created)
BT ~ 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 2 PRECURSOR.
GN THBS2 OR TSP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92147683; PubMed=1371115;
RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seidlin M.F.,
RA Dixit V.M.;
RT "Characterization of mouse thrombospondin 2 sequence and expression
during cell growth and development.";
RL J. Biol. Chem. 267:3274-3281(1992).
RN [2]
RP SEQUENCE OF 1-873 FROM N.A.
RX MEDLINE=91302287; PubMed=1712771;
RA Bornstein P., O'Rourke K., Wikstrom K., Wolf F.W., Katz R., Li P.,
RA Dixit V.M.;
RT "A second, expressed thrombospondin gene (Thbs2) exists in the mouse
genome.";
RL J. Biol. Chem. 266:12821-12824(1991).
CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
LAMININ AND TYPE V COLLAGEN.
CC -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 WMFC DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L07803; AAA53064.1; -;
DR EMBL; M64866; AAA40432.1; -;
DR PIR; A42587; A42587.
DR PIR; A39851; A39851.
DR HSP; P00740; 11XA.
DR MGD; MGI:98738; Thbs2.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; WMFC.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 9.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00092; TSP1; 3.
DR PROSITE; PS01208; WMFC; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1172
FT DOMAIN 19 232
FT DOMAIN 318 375
FT DOMAIN 381 432
FT DOMAIN 437 493
FT DOMAIN 494 548
FT DOMAIN 549 589
FT DOMAIN 590 647
FT DOMAIN 648 692
FT DOMAIN 725 760
FT DOMAIN 761 783
FT DOMAIN 784 819
FT DOMAIN 820 842
FT DOMAIN 843 880
FT DOMAIN 881 916
FT DOMAIN 917 952
FT DOMAIN 953 1172
FT SITE 928 930
FT DISULFID 266 266
FT DISULFID 270 270
FT DISULFID 553 564
FT DISULFID 558 574
FT DISULFID 577 588
FT DISULFID 594 610
FT DISULFID 601 619
FT DISULFID 622 646
FT DISULFID 652 665
FT DISULFID 659 678
FT DISULFID 680 691
FT CARBOHYD 151 151
FT CARBOHYD 316 316
FT CARBOHYD 330 330
FT CARBOHYD 457 457
FT CARBOHYD 584 584
FT CARBOHYD 710 710
FT CARBOHYD 1069 1069
SQ SEQUENCE 1172 AA; 129955 MW; 2AC7BB230E44C6F5 CRC64;
Query Match 100.0%; Score 84; DB 1; Length 1172;
Best Local Similarity 100.0%; Pred. No. 1.le-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 RESHFRGLQNHLVLF 16
|||||
Db 198 RESHFRGLQNHLVLF 213
RESULT 2
TSP2_MOUSE STANDARD; PRT; 1172 AA.
AC 003350;
DT 01-JUN-1994 (Rel. 29, Created)
BT ~ 01-JUN-1994 (Rel. 29, Last sequence update)

FT DOMAIN 381 432 TSP TYPE-1 1.
 FT DOMAIN 437 493 TSP TYPE-1 2.
 FT DOMAIN 494 548 TSP TYPE-1 3.
 FT DOMAIN 549 589 EGF-LIKE 1.
 FT DOMAIN 590 647 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 648 692 EGF-LIKE 3.
 FT DOMAIN 725 760 TSP TYPE-3 1.
 FT DOMAIN 761 783 TSP TYPE-3 2.
 FT DOMAIN 784 819 TSP TYPE-3 3.
 FT DOMAIN 820 842 TSP TYPE-3 4.
 FT DOMAIN 843 880 TSP TYPE-3 5.
 FT DOMAIN 881 916 TSP TYPE-3 6.
 FT DOMAIN 917 952 TSP TYPE-3 7.
 FT DOMAIN 953 1172 C-TERMINAL.
 FT SITE 928 930 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 266 266 INTERCHAIN (PROBABLE).
 FT DISULFID 270 270 INTERCHAIN (PROBABLE).
 FT DISULFID 553 564 BY SIMILARITY.
 FT DISULFID 558 574 BY SIMILARITY.
 FT DISULFID 577 588 BY SIMILARITY.
 FT DISULFID 594 610 BY SIMILARITY.
 FT DISULFID 601 619 BY SIMILARITY.
 FT DISULFID 622 645 BY SIMILARITY.
 FT DISULFID 652 665 BY SIMILARITY.
 FT DISULFID 659 678 BY SIMILARITY.
 FT DISULFID 680 691 BY SIMILARITY.
 FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 710 710 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1069 1069 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1172 AA; 129911 MW; 7CE8E4E8599822AB CRC64;

Query Match 100.0%; Score 84; DB 1; Length 1172;
 Best Local Similarity 100.0%; Pred. No. 1.1e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RESHFRLQLQNVHLVF 16
 DB 198 RESHFRLQLQNVHLVF 213
 |||||

RESULT 3
 TSP2_CHICK
 ID TSP2_CHICK STANDARD; PRT; 1178 AA.
 AC P35440;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE THROMBOSPONDIN 2 PRECURSOR.
 GN THBS2 OR TSP2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91217026; PubMed=2022631;
 RA Lawler J., Duquette M., Ferro P.;
 RT "Cloning and sequencing of chicken thrombospondin.;"
 RL J. Biol. Chem. 266:8039-8043(1991).
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 CC LAMININ AND TYPE V COLLAGEN.
 CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.

CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M60853; AAA51437.1; -;
 DR PIR; A39804; A39804.
 DR HSP; P00740; IIXA.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF-Ca.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR003367; VWFC.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF00090; tsp_1; 3.
 DR Pfam; PF02412; tsp_3; 8.
 DR Pfam; PF00093; vwc; 1.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00001; EGF_Like; 1.
 DR SMART; SM00209; TSP1; 3.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS00092; TSP1; 3.
 DR PROSITE; PS01208; VWFC; 1.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 ?
 FT CHAIN ? 1178 THROMBOSPONDIN 2.
 FT DOMAIN ? 232 HEPARIN-BINDING (POTENTIAL).
 FT DOMAIN 324 381 VWFC.
 FT DOMAIN 387 438 TSP TYPE-1 1.
 FT DOMAIN 441 499 TSP TYPE-1 2.
 FT DOMAIN 500 553 TSP TYPE-1 3.
 FT DOMAIN 555 595 EGF-LIKE 1.
 FT DOMAIN 596 653 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 654 698 EGF-LIKE 3.
 FT DOMAIN 731 766 TSP TYPE-3 1.
 FT DOMAIN 767 789 TSP TYPE-3 2.
 FT DOMAIN 790 825 TSP TYPE-3 3.
 FT DOMAIN 826 848 TSP TYPE-3 4.
 FT DOMAIN 849 886 TSP TYPE-3 5.
 FT DOMAIN 887 922 TSP TYPE-3 6.
 FT DOMAIN 923 958 TSP TYPE-3 7.
 FT DOMAIN 959 1178 C-TERMINAL.
 FT SITE 934 935 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 559 570 BY SIMILARITY.
 FT DISULFID 564 580 BY SIMILARITY.
 FT DISULFID 583 594 BY SIMILARITY.
 FT DISULFID 600 616 BY SIMILARITY.
 FT DISULFID 607 625 BY SIMILARITY.
 FT DISULFID 628 652 BY SIMILARITY.
 FT DISULFID 658 671 BY SIMILARITY.
 FT DISULFID 665 684 BY SIMILARITY.
 FT DISULFID 686 697 BY SIMILARITY.
 FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 590 590 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 716 716 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1075 1075 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1178 AA; 131816 MW; F37E02F42C8717A2 CRC64;


```
DR EMBL; M62470; AAA50611.1; -.
DR EMBL; M62450; AAA50611.1; JOINED.
DR EMBL; M62451; AAA50611.1; JOINED.
DR EMBL; M62452; AAA50611.1; JOINED.
DR EMBL; M62453; AAA50611.1; JOINED.
DR EMBL; M62454; AAA50611.1; JOINED.
DR EMBL; M62455; AAA50611.1; JOINED.
DR EMBL; M62456; AAA50611.1; JOINED.
DR EMBL; M62457; AAA50611.1; JOINED.
DR EMBL; M62458; AAA50611.1; JOINED.
DR EMBL; M62459; AAA50611.1; JOINED.
DR EMBL; M62460; AAA50611.1; JOINED.
DR EMBL; M62461; AAA50611.1; JOINED.
DR EMBL; M62462; AAA50611.1; JOINED.
DR EMBL; M62463; AAA50611.1; JOINED.
DR EMBL; M62464; AAA50611.1; JOINED.
DR EMBL; M62465; AAA50611.1; JOINED.
DR EMBL; M62466; AAA50611.1; JOINED.
DR EMBL; M62467; AAA50611.1; JOINED.
DR EMBL; M62468; AAA50611.1; JOINED.
DR EMBL; M62469; AAA50611.1; JOINED.
DR EMBL; M87276; AAA53063.1; -.
DR EMBL; J05606; AAA40431.1; -.
DR EMBL; J05605; AAA40431.1; JOINED.
DR PIR; A40558; A40558.
DR PIR; B42587; B42587.
DR PIR; A37905; A37905.
DR HSP; P35555; IEMO.
DR MGD; MGI:98737; Thbs1.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF-Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; VWFC.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 8.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00092; TSP1; 3.
DR PROSITE; PS01208; VWFC; 1.
DR Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat; EGF-like domain; Signal.
KW SIGNAL 1 18
FT CHAIN 19 1170
FT DOMAIN 19 232
FT DOMAIN 316 373
FT DOMAIN 379 430
FT DOMAIN 435 491
FT DOMAIN 492 548
FT DOMAIN 549 587
FT DOMAIN 588 645
FT DOMAIN 646 690
FT DOMAIN 723 758
FT DOMAIN 759 781
FT DOMAIN 782 817
FT DOMAIN 818 840
FT DOMAIN 841 878
FT DOMAIN 879 914
FT DOMAIN 915 950
FT DOMAIN 951 1170
FT SITE 926 928
FT DISULFID 270 270
FT DISULFID 274 274
FT DISULFID 551 562
BY SIMILARITY.

FT DISULFID 556 572
FT DISULFID 575 586
FT DISULFID 592 608
FT DISULFID 599 617
FT DISULFID 620 644
FT DISULFID 650 663
FT DISULFID 657 676
FT DISULFID 678 689
FT CARBOHYD 248 248
FT CARBOHYD 360 360
FT CARBOHYD 708 708
FT CARBOHYD 1067 1067
FT CONFLICT 1025 1025
SQ SEQUENCE 1170 AA; 129646 MW; 0443EA93615E7F06 CRC64;

Query Match 52.4%; Score 44; DB 1; Length 1170;
Best Local Similarity 61.5%; Pred. No. 9.2;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 4 HPRGLQLQNVHLVF 16
Db 207 NFGVLQNVRFVF 219

RESULT 9
YKYL_CAEEL STANDARD; PRT; 361 AA.
AC Q19910;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 40.9 KDA PROTEIN F30B5.4 IN CHROMOSOME IV.
GN F30B5.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Miller N., Bradshaw H.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE OKL38 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U42437; AAA83493.1; -.
CC WormPep; F30B5.4; CE04465.
KW Hypothetical protein.
SQ SEQUENCE 361 AA; 40896 MW; DD6553EA57D546EE CRC64;

Query Match 50.0%; Score 42; DB 1; Length 361;
Best Local Similarity 53.8%; Pred. No. 5.9;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RESHFRGLQLQNVH 13
Db 293 RESHFSTVFENAH 305

RESULT 10
YR23_MYCTU STANDARD; PRT; 397 AA.
ID YR23_MYCTU
AC O33228;
DT 30-MAY-2000 (Rel. 39, Created)
```

30-MAY-2000 (Rel. 39, Last sequence update)
 20-AUG-2001 (Rel. 40, Last annotation update)
 HYPOHETICAL 43.7 KDA PROTEIN RV2723.
 RV2723 OR MT2795 OR MTCY154.03.
 Mycobacterium tuberculosis.
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1773;
 [1]
 SEQUENCE FROM N.A.
 STRAIN-H37RV;
 MEDLINE=98295987; PubMed=9634230;
 Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence."
 Nature 393:537-544(1998).
 [2]
 SEQUENCE FROM N.A.
 STRAIN=CDC 1551 / Oshkosh;
 Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 Bishai W.;
 "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains."
 Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE TERC FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 EMBL: Z98209; CAB10899.1; -
 EMBL: AB007108; AAK47112.1; -
 TIGR: MT2795; -
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 1 21 POTENTIAL.
 FT TRANSMEM 39 59 POTENTIAL.
 FT TRANSMEM 76 96 POTENTIAL.
 FT TRANSMEM 103 123 POTENTIAL.
 FT TRANSMEM 124 144 POTENTIAL.
 FT TRANSMEM 194 214 POTENTIAL.
 FT TRANSMEM 219 239 POTENTIAL.
 FT TRANSMEM 255 275 POTENTIAL.
 FT TRANSMEM 301 321 POTENTIAL.
 FT CONFLICT 250 250 G -> S (IN REF. 2).
 SQ SEQUENCE 397 AA; 43679 MW; B85CFDD414B4C3E CRC64;
 Query Match 47.6%; Score 40; DB 1; Length 397;
 Best Local Similarity 80.0%; Pred. No. 15;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 7 GLLQNVLVLF 16
 |||||
 Db 222 GLTQNVLVLF 231
 RESULT 11

C723_ARATH
 ID C723_ARATH STANDARD; PRT; 501 AA.
 AC O65785; O9LTL9;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CYTOCHROME P450 71B3 (EC 1.14.-.-).
 GN CYP71B3 OR AT3G26220 OR MTC11.13.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA; TISSUE=Seedling;
 RX MEDLINE=98281573; PubMed=9620263;
 RA Mizutani M., Ward E., Ohta D.;
 RT "Cytochrome P450 superfamily in Arabidopsis thaliana: Isolation of
 cDNAs, differential expression, and RFLP mapping of multiple
 cytochromes P450."
 RT Plant Mol. Biol. 37:39-52(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20277480; PubMed=10819329;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
 features of the regions of 4,504,864 bp covered by sixty P1 and TAC
 clones."
 RL DNA Res. 7:131-135(2000).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 EMBL: D78602; BAA28534.1; -
 EMBL: AB024038; BAB02443.1; -
 DR InterPro; IPR001128; Cyt_P450.
 DR Pfam; PF00067; p450; 1.
 DR PRINTS; PR00463; EP450I.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Transmembrane; Heme; Multigene family.
 FT TRANSMEM 2 22 POTENTIAL.
 FT BINDING 445 445 HEME (BY SIMILARITY).
 FT CONFLICT 79 82 KEAA -> QETT (IN REF. 1).
 FT SEQUENCE 501 AA; 57663 MW; B4A48E39166FBE5 CRC64;
 SQ
 Query Match 47.6%; Score 40; DB 1; Length 501;
 Best Local Similarity 45.5%; Pred. No. 19;
 Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
 QY 4 HFRGLLNQVHL 14
 |::|::|
 Db 293 HLKGIQNIYL 303
 RESULT 12
 NEL2_MOUSE
 ID NEL2_MOUSE STANDARD; PRT; 816 AA.
 AC Q61220;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROTEIN KINASE C-BINDING PROTEIN NELL2 PRECURSOR (NEL-LIKE PROTEIN 2)
 DE (MEL91 PROTEIN).
 GN NELL2 OR MEL91.

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Elkins D.A., Rossi J.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP TSP N-TERMINAL DOMAIN.
 RX MEDLINE=98153258; PubMed=9480764;
 RA Beckmann G., Hanke J., Bork P., Reich J.;
 RT "Merging extracellular domains: fold prediction for laminin G-like
 RT and amino-terminal thrombospondin-like modules based on homology to
 RT pentraxins.";
 RL J. Mol. Biol. 275:725-730(1998).
 CC -|- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -|- SIMILARITY: CONTAINS 1 TSP N-TERMINAL DOMAIN.
 CC -|- SIMILARITY: CONTAINS 5 VWFC DOMAINS.
 CC -|- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U59230; AB02924.1; ALT_INIT.
 DR HGSP; P00740; IIXA.
 DR MGD; MGI:1858510; Nell2.
 DR InterPro; IPR000152; Asx hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR001007; VWFC.
 DR Pfam; PF00008; EGF; 5.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF00093; VWC; 2.
 DR SMART; SM00179; EGF_CA; 3.
 DR SMART; SM00001; EGF_Like; 3.
 DR SMART; SM00282; LamG; 1.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00214; VWC; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 3.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 3.
 DR PROSITE; PS01187; EGF_CA; 3.
 DR PROSITE; PS01208; VWFC; 2.
 KW Glycoprotein; EGF-like domain; Repeat; Signal.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 816 PROTEIN KINASE C-BINDING PROTEIN NELL2.
 FT DOMAIN 30 258 TSP N-TERMINAL.
 FT DOMAIN 272 331 VWFC 1.
 FT DOMAIN 332 396 VWFC 2.
 FT DOMAIN 397 439 EGF-LIKE 1.
 FT DOMAIN 440 481 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 482 522 EGF-LIKE 3.
 FT DOMAIN 521 553 EGF-LIKE 4.
 FT DOMAIN 555 601 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 602 637 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 638 693 VWFC 3.
 FT DOMAIN 698 756 VWFC 4.
 FT DOMAIN 758 813 VWFC 5.
 FT DISULFID 401 413 BY SIMILARITY.
 FT DISULFID 407 422 BY SIMILARITY.
 FT DISULFID 424 438 BY SIMILARITY.
 FT DISULFID 444 457 BY SIMILARITY.
 FT DISULFID 451 466 BY SIMILARITY.
 FT DISULFID 468 480 BY SIMILARITY.
 FT DISULFID 486 499 BY SIMILARITY.

FT DISULFID 493 508 BY SIMILARITY.
 FT DISULFID 510 521 BY SIMILARITY.
 FT DISULFID 525 535 BY SIMILARITY.
 FT DISULFID 529 541 BY SIMILARITY.
 FT DISULFID 543 552 BY SIMILARITY.
 FT DISULFID 559 572 BY SIMILARITY.
 FT DISULFID 566 581 BY SIMILARITY.
 FT DISULFID 583 600 BY SIMILARITY.
 FT DISULFID 606 619 BY SIMILARITY.
 FT DISULFID 613 628 BY SIMILARITY.
 FT CARBOHYD 630 636 BY SIMILARITY.
 FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 225 225 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 298 298 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 517 517 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 615 615 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 635 635 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 816 AA; 91163 MW; 5BDD0A946F87E74D CRC64;
 Query Match 47.6%; Score 40; DB 1; Length 816;
 Best Local Similarity 41.7%; Pred No. 31;
 Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;
 OY 4 HFRGLQNVHLV 15
 !:!:!:!:!
 DB 200 YFKGIMQDVHL 211
 RESULT 13
 SERA_ECOLI
 ID SERA_ECOLI STANDARD; PRT; 409 AA.
 AC P08328;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95) (PGDH).
 GN SERA OR B2913.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=86304370; PubMed=3017965;
 RA Tobey K.L., Grant G.A.;
 RT "The nucleotide sequence of the serA gene of Escherichia coli and the
 RT amino acid sequence of the encoded protein, D-3-phosphoglycerate
 RT dehydrogenase.";
 RL J. Biol. Chem. 261:12179-12183(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE OF 1-98 FROM N.A.
 RC STRAIN-K12;
 RA Roy I., Leadlay P.F.;
 RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 1-12.
 RC STRAIN-K12 / EMG2;
 RX MEDLINE=97443975; PubMed=9298646;
 RA Link A.J., Robison K., Church G.M.;
 RT "Comparing the predicted and observed properties of proteins encoded

in the genome of *Escherichia coli* K-12.";
Electrophoresis 18:1259-1313(1997).
[5]
X-RAY CRYSTALLOGRAPHY (2.75 ANGSTROMS).
MEDLINE-95236205; PubMed-7719856;
Schuller D.J., Grant G.A., Banaszak L.J.;
"The allosteric ligand site in the Vmax-type cooperative enzyme
phosphoglycerate dehydrogenase.";
Nat. Struct. Biol. 2:69-76(1995).
-1- CATALYTIC ACTIVITY: 3-PHOSPHOGLYCERATE + NAD(+) =
3-PHOSPHOHYDROXYPYRUVATE + NADH.
-1- ENZYME REGULATION: IN BACTERIA DISPLAYS FEEDBACK INHIBITION BY
L-SERINE.
-1- PATHWAY: FIRST COMMITTED STEP IN THE "PHOSPHORYLATED" PATHWAY
OF L-SERINE BIOSYNTHESIS.
-1- SUBUNIT: HOMOTETRAMER.
-1- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID
DEHYDROGENASES FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X66836; CAA47308.1; -;
DR EMBL; U28377; AAA69080.1; -;
DR EMBL; AE000374; AAC75950.1; -;
DR EMBL; L29397; AAA24625.1; -;
DR EMBL; A25268; CAA01762.1; -;
DR PIR; A25200; DEECPC;
DR PIR; S22096; S22096;
DR PDB; IPSD; 31-JUL-95.
DR ECO2DBASE; G044.1; 6TH EDITION.
DR EcoGene; EGI0944; serA.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR002162; D_2_hydroxyacid_DH.
DR Pfam; PF00389; 2-Hacid_DH; 1.
DR Pfam; PF01842; ACT; 1.
DR PROSITE; PS00065; D_2-HYDROXYACID_DH_1; 1.
DR PROSITE; PS00670; D_2-HYDROXYACID_DH_2; 1.
DR PROSITE; PS00671; D_2-HYDROXYACID_DH_3; 1.
Serine biosynthesis; Oxidoreductase; NAD; 3D-structure;
Complete proteomes.
KW INIT_MET 0
FT NP_BIND 152 180 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 239 239 SUBSTRATE-BINDING (BY SIMILARITY).
FT ACT_SITE 268 268 BY SIMILARITY.
FT ACT_SITE 291 291 BY SIMILARITY.
SQ SEQUENCE 409 AA; 44044 MW; F15F8AFB86FA1D6B CRC64;

Query Match 46.4%; Score 39; DB 1; Length 409;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 RESHFRGLQNVL 14
Db 51 RDAHFGLRSRTHL 64
|::||| ||
|::||| ||

RESULT 14
DYHC_ANTCR STANDARD; PRT; 4466 AA.
AC P39057;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DYNEIN BETA CHAIN, CILIARY.
OS Anthocidaris crassispina (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

QY 1 RESHFRGLQNVL 15
Db 3931 KEGHW-VILQNIHLV 3944
|:| |::||| |||
|:| |::||| |||

RESULT 15
DYHC_TRIGR STANDARD; PRT; 4466 AA.
AC P23098;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DYNEIN BETA CHAIN, CILIARY.
OS Tripneustes gratilla (Hawaiian sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Temnopleuroidea; Toxopneustidae;
OC Tripneustes
OX NCBI_TaxID=7673;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Blastula;
RX MEDLINE-91326103; PubMed-1830927;

OC Echinoidea; Euechinoidea; Echinacea; Echinomastixidae;
OX Anthocidaris
OX NCBI_TaxID=7629;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91326104; PubMed-1830928;
RA Ogawa K.;
RT "Four ATP-binding sites in the midregion of the beta heavy chain of
dynein.";
RL Nature 352:643-645(1991).
CC -1- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND
CC FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.
CC DYNEIN HAS ATPASE ACTIVITY.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS (ALPHA AND BETA),
CC THREE INTERMEDIATE CHAINS AND SEVERAL LIGHT CHAINS.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D01021; BAA00827.1; -;
DR PIR; S17231; S17231.
KW Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
KW Coiled coil.
FT DOMAIN 482 502 COILED COIL (POTENTIAL).
FT DOMAIN 627 643 COILED COIL (POTENTIAL).
FT DOMAIN 734 805 COILED COIL (POTENTIAL).
FT DOMAIN 1036 1056 COILED COIL (POTENTIAL).
FT DOMAIN 1306 1337 COILED COIL (POTENTIAL).
FT DOMAIN 1443 1468 COILED COIL (POTENTIAL).
FT DOMAIN 1950 1978 MICROTUBULE-BINDING (POTENTIAL).
FT DOMAIN 3033 3134 COILED COIL (POTENTIAL).
FT DOMAIN 3263 3325 COILED COIL (POTENTIAL).
FT DOMAIN 3573 3642 COILED COIL (POTENTIAL).
FT NP_BIND 154 161 ATP (POTENTIAL).
FT NP_BIND 1852 1859 ATP (POTENTIAL).
FT NP_BIND 2133 2140 ATP (POTENTIAL).
FT NP_BIND 2460 2467 ATP (POTENTIAL).
FT NP_BIND 2805 2812 ATP (POTENTIAL).
SQ SEQUENCE 4466 AA; 511772 MW; C465CC5C6D325D CRC64;

Query Match 45.8%; Score 38.5; DB 1; Length 4466;
Best Local Similarity 53.3%; Pred. No. 3.3e+02;
Matches 8; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

QY 1 RESHFRGLQNVL 15
Db 3931 KEGHW-VILQNIHLV 3944
|:| |::||| |||
|:| |::||| |||

RESULT 15
DYHC_TRIGR STANDARD; PRT; 4466 AA.
AC P23098;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DYNEIN BETA CHAIN, CILIARY.
OS Tripneustes gratilla (Hawaiian sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Temnopleuroidea; Toxopneustidae;
OC Tripneustes
OX NCBI_TaxID=7673;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Blastula;
RX MEDLINE-91326103; PubMed-1830927;

```

RA Gibbons I.R., Gibbons B.H., Mocz G., Asai D.J.;
RT "Multiple nucleotide-binding sites in the sequence of dynein beta
RN heavy chain.";
RL Nature 352:640-643(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92020893; PubMed=1833761;
RA Gibbons I.R., Asai D.J., Ching N.S., Dolecki G.J., Mocz G.,
RT Philipson C.A., Ren H., Tang W.Y., Gibbons B.H.;
RA "A PCR procedure to determine the sequence of large polypeptides by
RT rapid walking through a cDNA library.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:8563-8567(1991).
CC -!- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND
CC FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.
CC DYNEIN HAS ATPASE ACTIVITY.
CC -!- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS (ALPHA AND BETA),
CC THREE INTERMEDIATE CHAINS AND SEVERAL LIGHT CHAINS.
CC -!- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59603; CAA42170.1; -.
DR PIR; S17653; S17653.
KW Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
KW Coiled coil.
FT DOMAIN 627 643 COILED COIL (POTENTIAL).
FT DOMAIN 733 805 COILED COIL (POTENTIAL).
FT DOMAIN 1036 1056 COILED COIL (POTENTIAL).
FT DOMAIN 1306 1337 COILED COIL (POTENTIAL).
FT DOMAIN 1443 1468 COILED COIL (POTENTIAL).
FT DOMAIN 1850 1878 MICROTUBULE-BINDING (POTENTIAL).
FT DOMAIN 3033 3092 COILED COIL (POTENTIAL).
FT DOMAIN 3263 3325 COILED COIL (POTENTIAL).
FT DOMAIN 3573 3642 COILED COIL (POTENTIAL).
FT NP_BIND 154 161 ATP (POTENTIAL).
FT NP_BIND 1852 1859 ATP (POTENTIAL).
FT NP_BIND 2133 2140 ATP (POTENTIAL).
FT NP_BIND 2460 2467 ATP (POTENTIAL).
FT NP_BIND 2805 2812 ATP (POTENTIAL).
FT VARIANT 611 615 MISSING.
FT VARIANT 3356 3358 LPG -> LLTGNEFFCCFMTAG.
SQ SEQUENCE 4466 AA; 511771 MW; 2A695BF8F336911E CRC64;

```

```

Query Match 45.8%; Score 38.5; DB 1; Length 4466;
Best Local Similarity 53.3%; Pred. No. 3.3e+02;
Matches 8; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

```

```

Qy 1 RESHFRGLQNVHLV 15
   :| :| :| :| :|
Db 3931 KEGHW-VILQNIHLV 3944

```

```

Search completed: January 31, 2002, 09:36:48
Job time: 767 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:36:18 ; Search time 90.11 Seconds
(without alignments)
11.363 Million cell updates/sec

Title: US-09-536-087-10
Perfect score: 53
Sequence: 1 WSPNAEW 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	94.3	802	5 Q19764	Q19764 caenorhabdi
2	50	94.3	1202	4 Q9P283	Q9P283 homo sapien
3	48	90.6	788	5 Q22631	Q22631 caenorhabdi
4	46	86.8	444	5 Q9VWA7	Q9VWA7 drosophila
5	45	84.9	215	13 Q93392	Q93392 coturnix co
6	45	84.9	685	6 Q9RTS5	Q9RTS5 bos taurus
7	45	84.9	945	11 Q08722	Q08722 rattus norv
8	45	84.9	945	11 Q9D398	Q9D398 mus musculu
9	45	84.9	1081	5 Q9V631	Q9V631 drosophila
10	45	84.9	1083	5 Q9VTT0	Q9VTT0 drosophila
11	44	83.0	408	2 Q44365	Q44365 agrobacteri
12	44	83.0	425	6 Q02661	Q02661 bos taurus
13	44	83.0	1184	4 Q95339	Q95339 homo sapien
14	43	81.1	931	4 Q95185	Q95185 homo sapien
15	43	81.1	1235	4 Q95428	Q95428 homo sapien
16	42	79.2	213	2 Q9XC55	Q9XC55 streptomyce
17	42	79.2	654	5 Q19284	Q19284 caenorhabdi
18	41	77.4	244	11 Q9CRC7	Q9CRC7 mus musculu
19	41	77.4	415	2 Q9R6E8	Q9R6E8 agrobacteri

20	41	77.4	571	2 Q9CMT4	Q9CMT4 pasteurella
21	41	77.4	708	10 Q9FZG1	Q9FZG1 arabidopsis
22	41	77.4	736	5 Q93243	Q93243 caenorhabdi
23	41	77.4	898	5 Q76822	Q76822 branchiosto
24	41	77.4	898	11 Q08721	Q08721 rattus norv
25	41	77.4	931	11 Q08747	Q08747 mus musculu
26	41	77.4	957	5 Q19204	Q19204 caenorhabdi
27	41	77.4	1444	5 Q17591	Q17591 caenorhabdi
28	41	77.4	1637	6 Q9XSV8	Q9XSV8 bos taurus
29	41	77.4	4123	4 Q75851	Q75851 homo sapien
30	40	75.5	169	2 Q9RZC9	Q9RZC9 deinococcus
31	40	75.5	702	10 Q9M0S5	Q9M0S5 arabidopsis
32	40	75.5	870	6 Q02660	Q02660 bos taurus
33	39	73.6	142	2 Q9RRN1	Q9RRN1 deinococcus
34	39	73.6	147	11 Q61427	Q61427 mus musculu
35	39	73.6	165	10 Q9SNN3	Q9SNN3 oryza sativ
36	39	73.6	175	4 Q9PLJ8	Q9PLJ8 homo sapien
37	39	73.6	206	4 Q9NQ91	Q9NQ91 homo sapien
38	39	73.6	238	5 Q76510	Q76510 cryptospori
39	39	73.6	256	2 Q9F5M2	Q9F5M2 rhodocyclus
40	39	73.6	335	3 Q9Y800	Q9Y800 schizosacch
41	39	73.6	359	7 Q9BCY6	Q9BCY6 papio cynoc
42	39	73.6	383	5 Q76138	Q76138 trypanosoma
43	39	73.6	387	2 Q9F3D8	Q9F3D8 agrobacteri
44	39	73.6	389	6 Q97887	Q97887 bos taurus
45	39	73.6	417	11 Q9ESP8	Q9ESP8 rattus norv

ALIGNMENTS

RESULT	1
Q19764	
ID	Q19764 PRELIMINARY; PRT; 802 AA.
AC	Q19764;
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	F23H12.5 PROTEIN.
GN	F23H12.5
OS	Caenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC	Rhabditidae; Peloderinae; Caenorhabditis.
OX	NCBI_TaxID=6239;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Kershaw J.;
RL	Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=94150718; PubMed=7906398;
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA	Craxton M., Dear S., Du Z., Durbin R., Favello A., Jier M., Johnston L.,
RA	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA	Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA	Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA	Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA	Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA	Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA	Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL	elegans."
RL	Nature 368:32-38(1994).
DR	EMBL: Z74472; CAA98943.1;
DR	InterPro: IPR000884; TSP1.
DR	SMART: PSS0092; TSP1; 2.
DR	SMART: SM00209; TSP1; 2.
SQ	SEQUENCE 802 AA; 90137 MW; 0A9CA26F60A62EE5 CRC64;

Query Match 94.3% Score 50; DB 5; Length 802;
Best Local Similarity 85.7%; Pred. No. 5.9;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7

||||:||

Db 706 WSPWSEW 712

RESULT 2

ID Q9P283 PRELIMINARY; PRT; 1202 AA.

AC Q9P283;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE KIAA1445 PROTEIN (FRAGMENT).

GN KIAA1445.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RP [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20277482; PubMed=10819331;

RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;

RT "Prediction of the coding sequences of unidentified human

genes.XVII.The complete sequences of 100 new cDNA clones from brain

which code for large proteins in vitro."

RL DNA Res. 7:143-150(2000).

DR EMBL; AB040878; BAA5969.1;

DR InterPro; IPR002165; Plexin_repeat.

DR InterPro; IPR003659; PSI.

DR InterPro; IPR001627; Sema.

DR InterPro; IPR000884; TSP1.

DR Pfam; PF01437; plexin_repeat; 1.

DR Pfam; PF01403; Sema; 1.

DR Pfam; PF00090; tsp_1; 5.

DR SMART; SM00423; PSI; 1.

DR SMART; SM00209; TSP1; 4.

DR PROSITE; PS50092; TSP1; 2.

FT NON_TER 1

SQ SEQUENCE 1202 AA; 131154 MW; 1E1FE9580953484 CRC64;

Query Match

Best Local Similarity 94.3%; Score 50; DB 4; Length 1202;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7

||||:||

Db 1019 WSPWSEW 1025

RESULT 3

ID Q22631

AC Q22631 PRELIMINARY; PRT; 788 AA.

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE T21B6.3 PROTEIN.

GN T21B6.3.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RP [1]

RP SEQUENCE FROM N.A.

RA Cottage A.;

RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z68011; CRA92014.1;
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR002965; P-rich_extensn.
DR InterPro; IPR003609; Pan_app.
DR Pfam; PF00090; tsp_1; 6.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00473; PAN_AP; 1.
DR SMART; SM00209; TSP1; 7.
SQ SEQUENCE 788 AA; 88727 MW; BB0DD9F1D29BD961 CRC64;

Query Match 90.6%; Score 48; DB 5; Length 788;

Best Local Similarity 85.7%; Pred. No. 11;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7

||||:||

Db 606 WSPWQEW 612

RESULT 4

Q9VWA7

ID Q9VWA7 PRELIMINARY; PRT; 444 AA.

AC Q9VWA7;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CGI3817 PROTEIN.

GN CGI3817.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=72727;

RP [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,

RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jaisi M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003514; AAF49041.1; -;
 DR FlyBase; FBgn0036963; CG13817.
 DR InterPro; IPR000884; TSP1.
 DR Pfam; PF00090; tsp_1; 1.
 DR SMART; SM00209; TSP1; 1.
 DR PROSITE; PS50092; TSP1; 1.
 SQ SEQUENCE 444 AA; 49690 MW; E05ADB9AEE78A385 CRC64;

Query Match 86.8%; Score 46; DB 5; Length 444;
 Best Local Similarity 71.4%; Pred. No. 13;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
 | |||
 Db 402 WGPWSEW 408

RESULT 5
 O93392 PRELIMINARY; PRT; 215 AA.
 AC O93392;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE PRO-ALPHA2(1) COLLAGEN (FRAGMENT).
 OS Coturnix coturnix japonica (Japanese quail).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Coturnix.
 OX NCBI_TaxID=93934;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Oberst C., Hartl M., Weiskirchen R., Bister K.;
 RT "Conditional cell transformation by doxycycline-controlled expression
 RT of the MC29 v-myc allele."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Weiskirchen R., Bister K.;
 RT "Isolation of quail pro-alpha2(1) collagen mRNA."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF077329; AAD12181.1; -;
 DR InterPro; IPR000087; Collagen.
 DR Pfam; PF01391; Collagen; 3.
 FT NON_TER 1
 SQ SEQUENCE 215 AA; 20194 MW; CDE607AE07D03126 CRC64;

Query Match 84.9%; Score 45; DB 13; Length 215;
 Best Local Similarity 71.4%; Pred. No. 9;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
 |||||
 Db 206 WSPWQW 212

RESULT 6
 O9TTSS PRELIMINARY; PRT; 685 AA.
 AC O9TTSS;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE SCO-SPONDIN (FRAGMENT).
 GN SCO-SPONDIN.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-SUBCOMMISSURAL ORGAN;
 RA Gobron S., Creveaux I., Monnerie H., Elbitar F.;
 RT "Characterization of cattle SCO-spondin."
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ132106; CAB53759.1; -;
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR001007; VWFC.
 DR InterPro; IPR002919; TIL.
 DR Pfam; PF00090; tsp_1; 6.
 DR Pfam; PF01826; TIL; 2.
 DR PROSITE; PS50092; TSP1; 4.
 DR SMART; SM00209; TSP1; 6.
 DR SMART; SM00011; VWC_def; 2.
 DR SMART; SM00215; VWC_out; 1.
 FT NON_TER 1
 SQ SEQUENCE 685 AA; 71389 MW; A003F880E746D5D3 CRC64;

Query Match 84.9%; Score 45; DB 6; Length 685;
 Best Local Similarity 71.4%; Pred. No. 28;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
 | |||
 Db 494 WGPWSEW 500

RESULT 7
 O08722 PRELIMINARY; PRT; 945 AA.
 AC O08722;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE TRANSMEMBRANE RECEPTOR UNC5H2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-97271897; PubMed-9126742;
 RA Leonardo E.D., Hinck L., Masu M., Keino-Masu K., Ackerman S.L.,
 RA Tessier-Lavigne M.;
 RT "Vertebrate homologues of *C. elegans* UNC-5 are candidate netrin
 RT receptors."
 RL Nature 386:833-837(1997).
 CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.
 DR EMBL; U87306; AAB57679.1; -;
 DR InterPro; IPR000488; Death.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR000906; Z05.
 DR Pfam; PF00531; death; 1.

```
DR Pfam; PF00047; ig; 1.
DR Pfam; PF00090; tsp_1; 2.
DR Pfam; PF00791; ZU5; 1.
DR SMART; SM00005; DEATH; 1.
DR SMART; SM00408; IGC2; 1.
DR SMART; SM00209; TSP1; 2.
DR SMART; SM00218; ZU5; 1.
DR PROSITE; PS00209; TSP1; 2.
DR SMART; SM00218; ZU5; 1.
DR PROSITE; PS0017; DEATH_DOMAIN; 1.
DR PROSITE; PS0092; TSP1; 2.
KW Transmembrane.
SQ SEQUENCE 945 AA; 103520 MW; 6E9C2A262E560B9B CRC64;

Query Match      84.9%; Score 45; DB 11; Length 945;
Best Local Similarity 85.7%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
DB 249 WSSWAEW 255

RESULT 8
QD398 PRELIMINARY; PRT; 945 AA.
AC QD398;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DE 6330415E02RIK PROTEIN.
GN 6330415E02RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=MEDULLA OBLONGATA;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Nombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL; AK018177; BAB31108.1; -.
DR MGD; MGI:1924103; 6330415E02RIK.
DR InterPro; IPR000488; Death.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000906; ZU5.
DR Pfam; PF00531; death; 1.
DR Pfam; PF00047; ig; 1.
DR Pfam; PF00090; tsp_1; 2.
DR Pfam; PF00791; ZU5; 1.
```

```
DR SMART; SM00005; DEATH; 1.
DR SMART; SM00409; IG; 2.
DR SMART; SM00408; IGC2; 1.
DR SMART; SM00209; TSP1; 2.
DR SMART; SM00218; ZU5; 1.
DR PROSITE; PS0017; DEATH_DOMAIN; 1.
DR PROSITE; PS0092; TSP1; 2.
SQ SEQUENCE 945 AA; 103725 MW; 43D33B4524E0CBF2 CRC64;

Query Match      84.9%; Score 45; DB 11; Length 945;
Best Local Similarity 85.7%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
DB 249 WSSWAEW 255

RESULT 9
Q9U631 PRELIMINARY; PRT; 1081 AA.
AC Q9U631;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE SEMAPHORIN 5C PROTEIN.
GN SEMA-5C OR CG5661.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RA Khare N., Daroch S., Chiquet-Ehrismann R., Baumgartner S.;
RT "Expression patterns of two new members of the semaphorin family in
RT Drosophila suggest early functions during embryogenesis.";
RL Mech. Dev. 0:0-0(1999).
DR EMBL; AF198084; AAF04860.1; -.
DR FlyBase; FBgn028679; SEMA-5C.
DR InterPro; IPR000183; Orn_DAP_Arg_decarboxylase.
DR InterPro; IPR002165; Plexin_repeat.
DR InterPro; IPR003659; PSI.
DR InterPro; IPR001627; SEMA.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF01437; Plexin_repeat; 1.
DR Pfam; PF01403; SEMA; 1.
DR Pfam; PF00090; tsp_1; 6.
DR SMART; SM00423; PSI; 1.
DR SMART; SM00209; TSP1; 6.
DR PROSITE; PS00879; QDR_DC_2; UNKNOWN_1.
DR PROSITE; PS0092; TSP1; 8.
SQ SEQUENCE 1081 AA; 120435 MW; 561071C831C431D3 CRC64;

Query Match      84.9%; Score 45; DB 5; Length 1081;
Best Local Similarity 71.4%; Pred. No. 44;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
DB 662 WGPWGEW 668

RESULT 10
Q9VTT0 PRELIMINARY; PRT; 1083 AA.
AC Q9VTT0;
ID Q9VTT0;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
```

DE SEMA-5C PROTEIN.
GN SEMA-5C OR CG5661.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophiliidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Arif J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Brodeur E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleeb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "the genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003542; AAF49966.2;
DR FlyBase; FBgn028679; Sema-5c.
DR InterPro; IPR000183; Orn.DAP.Arg.decarboxylase.
DR InterPro; IPR002165; Plectin_repeat.
DR InterPro; IPR003659; PSI.
DR InterPro; IPR001627; Sema.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF01437; Plectin_repeat; 2.
DR Pfam; PF01403; Sema; 1.
DR Pfam; PF00090; tsp_1; 7.
DR SMART; SM00423; PSI; 1.
DR SMART; SM00209; TSPI; 6.
DR PROSITE; PS00879; ODR_DC_2_2; UNKNOWN_1.
DR PROSITE; PS50092; TSPI; 4.
SQ SEQUENCE 1083 AA; 120744 MW; 102885FC56BDBDEA CRC64;

Query Match 84.9%; Score 45; DB 5; Length 1083;
Best Local Similarity 71.4%; Pred. No. 44;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
| | | | |
Db 664 WGPWGEW 670

RESULT 11
Q44365 ID Q44365 PRELIMINARY; PRT; 408 AA.
AC Q44365; 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE TRAB PROTEIN.
GN TRAB.
OS Agrobacterium radiobacter.
OG Plasmid Ti.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=358;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96312368; PubMed=8763954;
RA Alt-Morbe J., Stryker J.L., Fuqua C., Li P.L., Farrand S.K.,
RA Winans S.C.;
RT "The conjugal transfer system of Agrobacterium tumefaciens octopine-
RT type Ti plasmids is closely related to the transfer system of an IncP
RT plasmid and distantly related to Ti plasmid vir genes.";
RL J. Bacteriol. 178:4248-4257(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX Winans S.C., Zhu J., Oger P.M., Schrammeijer B., Hooykaas P.J.,
RA Farrand S.K.;
RT "Octopine-type Ti plasmid sequence.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF242881; AAF71137.1; -.
KW Plasmid.
SQ SEQUENCE 408 AA; 44108 MW; 38B83993756CDB60 CRC64;

Query Match 83.0%; Score 44; DB 2; Length 408;
Best Local Similarity 71.4%; Pred. No. 24;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
| | | | |
Db 314 WOPWLEW 320

RESULT 12
Q02661 ID Q02661 PRELIMINARY; PRT; 425 AA.
AC Q02661;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SCO-SPONDIN (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SUBCOMMISSURAL ORGAN;
RA Gobron S.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y08561; CAA69868.1; -.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR001007; VWFC.
DR InterPro; IPR002919; TIL.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF01826; TIL_1.
DR PROSITE; PS50092; TSPI; 4.
DR SMART; SM00209; TSPI; 3.
DR SMART; SM00011; VWC_def; 1.
DR SMART; SM00215; VWC_out; 1.

```

FT NON_TER 1 1
SQ NON_TER 425 425
SQ SEQUENCE 425 AA; 43721 MW; E63DD4AFC6DB395D CRC64;

Query Match 83.0%; Score 44; DB 6; Length 425;
Best Local Similarity 71.4%; Pred. No. 25;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
DB 193 WTPWAPW 199
I:|||||

RESULT 13
ID 075339 PRELIMINARY; PRT; 1184 AA.
AC 075339;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CARTILAGE INTERMEDIATE LAYER PROTEIN.
GN CILP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ARTICULAR CARTILAGE;
RA MEDLINE=98389785; PubMed=9722584;
RX Lorenzo P., Neame P., Sommarin Y., Heinegard D.;
RT "Cloning and deduced amino acid sequence of a novel cartilage protein
RT (CILP) identifies a proform including a nucleotide
RT pyrophosphohydrolase."
RL J Biol. Chem. 273:23469-23475(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Nakamura I., Okawa A., Ikegawa S., Takaoka K., Nakamura Y.;
RT "Genomic organization, mapping, and polymorphisms of the gene encoding
RT human cartilage intermediate layer protein (CILP).";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Lorenzo P., Aman P., Sommarin Y., Heinegard D.;
RT "Pro-CILP: Gene structure and chromosomal localization.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL; AF035408; AAC33838.1; -.
DR EMBL; AB022430; BAA76692.1; -.
DR EMBL; AF035455; AAF14689.1; -.
DR EMBL; AF035448; AAF14689.1; JOINED.
DR EMBL; AF035449; AAF14689.1; JOINED.
DR EMBL; AF035451; AAF14689.1; JOINED.
DR EMBL; AF035453; AAF14689.1; JOINED.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR InterPro; IPR001451; Hexapep_transf.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00408; IGC2; 1.
DR SMART; SM00209; TSP1; 2.
DR SMART; SM00218; ZU5; 1.
DR PROSITE; PS50092; TSP1; 2.
DR Receptor; Transmembrane.
KW Receptor; Transmembrane.
SQ SEQUENCE 931 AA; 103101 MW; EFD71123C98DABB8 CRC64;

Query Match 81.1%; Score 43; DB 4; Length 931;
Best Local Similarity 57.1%; Pred. No. 74;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
DB 319 WTPWSKW 325
I:|||||

RESULT 15
ID 095428 PRELIMINARY; PRT; 1235 AA.
AC 095428;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE HYPOTHETICAL 133.5 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowan L., Madan A., Qian S., Abbasi N., Dors M., Ratcliffe A.,
RA Madan A., Dickhoff R., Shaffer T., James R., Lasky S., Hood L.;

```


RT *Complete sequence of the gene for presenilin 1.*;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.
 CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL; AF109907; AAC97963.1; -.
 DR HSSP; P12111; 2KNT.
 DR InterPro; IPR003598; Ig_C2.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR002223; Kunitz_BPTI.
 DR InterPro; IPR000884; TSPI.
 DR Pfam; PF00047; ig; 3.
 DR Pfam; PF00014; Kunitz_BPTI; 1.
 DR Pfam; PF00090; tsp_1; 5.
 DR PRINTS; PR00759; BASICPTASE.
 DR SMART; SM00408; IGC2; 3.
 DR SMART; SM00131; KU; 1.
 DR SMART; SM00209; TSPI; 5.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
 DR PROSITE; PS0279; BPTI_KUNITZ_2; 2.
 DR PROSITE; PS0092; TSPI; 4.
 KW Hypothetical protein; Serine protease inhibitor.
 SQ SEQUENCE 1235 AA; 133476 MW; A0B44CCE4F38E350 CRC64;

Query Match 81.1%; Score 43; DB 4; Length 1235;
 Best Local Similarity 57.1%; Pred. No. 98;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
 DB 29 WGPWSQW 35

Search completed: January 31, 2002, 09:36:19
 Job time: 788 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:20:35 ; Search time 49.78 Seconds
(without alignments)
24.484 Million cell updates/sec

Title: US-09-536-087-6
Perfect score: 84
Sequence: 1 RESHFRGLQNVHLVF 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84	100.0	1172	1 TSHUP2	thrombospondin 2 p
2	84	100.0	1172	2 A42587	thrombospondin 2 p
3	79	94.0	1178	1 A39804	thrombospondin pre
4	44	52.4	229	2 S57957	thrombospondin 1 -
5	44	52.4	1170	1 TSHUP1	thrombospondin 1 p
6	44	52.4	1170	2 A40558	thrombospondin 1 p
7	43	51.2	350	2 T46997	hypothetical prote
8	43	51.2	1383	2 T06091	hypothetical prote
9	42	50.0	361	2 T29571	hypothetical prote
10	41	48.8	177	2 F84245	inosine-5'-monopho
11	41	48.8	452	2 T28883	hypothetical prote
12	41	48.8	693	2 C86364	hypothetical prote
13	40	47.6	397	2 B70505	hypothetical prote
14	40	47.6	501	2 T52170	cytochrome P450 mo
15	40	47.6	614	2 T29902	hypothetical prote
16	40	47.6	752	2 T48574	hypothetical prote
17	39	46.4	201	2 E72321	hypothetical prote
18	39	46.4	410	1 DEECPG	phosphoglycerate d
19	39	46.4	410	2 D85947	hypothetical prote
20	39	46.4	865	2 T40288	conserved hypotet
21	39	46.4	1245	2 H83574	hypothetical prote
22	39	46.4	1286	2 B71413	hypothetical prote
23	38.5	45.8	136	2 T33423	hypothetical prote
24	38.5	45.8	4466	1 S17231	dynein beta heavy
25	38.5	45.8	4466	1 S17653	dynein beta heavy
26	38	45.2	249	2 H72200	conserved hypotet
27	38	45.2	351	2 H72328	cell division prot
28	38	45.2	622	2 S56214	probable membrane
29	38	45.2	750	1 JDVLVH	DNA-directed DNA p

protein kinase (EC
DNA-directed DNA p
protein kinase (EC
hypothetical prote
Bloom's syndrome r
probable multi-dom
herc2 protein - mo
probable membrane
hypothetical prote
minor tail protein
phosphoribosylglyc
phosphoribosylglyc
phosphoribosylglyc
hypothetical prote
thiamin biosynthet
thiamin biosynthes

ALIGNMENTS

RESULT 1

TSHUP2

thrombospondin 2 precursor - human

C:Species: Homo sapiens (man)

C:Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999

C:Accession: A47379; A42173

R:LaBell, T.L.; Byers, P.H.

Genomics 17, 225-229, 1993

A:Title: Sequence and characterization of the complete human thrombospondin 2 cDNA: p

A:Reference number: A47379; MUID:94010892

A:Accession: A47379

A:Molecule type: mRNA

A:Residues: 1-1172 <LAB>

A:Cross-references: GB:112350; NID:g307505; PIDN:AAA03703.1; PID:g307506

R:LaBell, T.L.; Milewicz, D.J.; Distche, C.M.; Byers, P.H.

Genomics 12, 421-429, 1992

A:Title: Thrombospondin II: partial cDNA sequence, chromosome location, and expressio

A:Reference number: A42173; MUID:92217961

A:Accession: A42173

A:Molecule type: mRNA

A:Residues: 560-1172 <LA2>

A:Cross-references: GB:M81339

A:Experimental source: fibroblast

A:Note: sequence extracted from NCBI backbone (NCBIN:95091, NCBI:P:95096)

C:Genetics:

A:Gene: GDB:THBS2; TSP2

A:Cross-references: GDB:128789; OMIM:188061

A:Map position: 6q27-6q27

C:Complex: homotrimer, disulfide linked

C:Function:

A:Description: participates in cell migration and adhesion, and in platelet aggregati

C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;

C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tri

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-1172/Product: thrombospondin 2 #status predicted <MAT>

F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>

F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>

F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>

F:493-549/Domain: thrombospondin type 1 repeat homology <EGF1>

F:652-691/Domain: EGF homology <EGF>

F:928-930/Region: cell attachment (R-G-D) motif

F:151,316,330,457,584,710,1069/Binding site: carbohydrate (Asn) (covalent) #status pr

F:167-226/Disulfide bonds: #status predicted

F:266,270/Disulfide bonds: interchain #status predicted

F:612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 100.0%; Score 84; DB 1; Length 1172;

Best Local Similarity 100.0%; Pred. No. 1.2e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: January 31, 2002, 09:24:52 ; Search time 49.78 Seconds
(without alignments)
27.544 Million cell updates/sec

Title: US-09-536-087-7
Perfect score: 110
Sequence: 1 PATCANP5FVEGECPCSC 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	100.0	1172	2 A42587	thrombospondin 2 p
2	105	95.5	1172	1 TSHUP2	thrombospondin 2 p
3	90	81.8	1178	1 A39804	thrombospondin pre
4	66	60.0	884	2 T18449	hypothetical prote
5	62	56.4	1497	2 I49607	procollagen type V
6	59	53.6	206	2 S18250	collagen alpha 1(I
7	59	53.6	615	2 A05269	collagen alpha 1(I
8	59	53.6	886	2 I50694	collagen alpha 1(I
9	59	53.6	1170	1 TSHUP1	thrombospondin 1 p
10	59	53.6	1170	2 A40558	thrombospondin 1 p
11	59	53.6	1376	2 S63986	collagen alpha 5 c
12	59	53.6	1487	2 B41182	collagen alpha 1(I
13	58.5	53.2	429	2 T21113	hypothetical prote
14	57	51.8	1496	1 CGHU2V	collagen alpha 2(V
15	56	50.9	98	2 I49562	alpha-1 type III c
16	56	50.9	1464	2 S59856	collagen alpha 1(I
17	55.5	50.5	776	2 S28258	androgen-regulated
18	55	50.0	1038	2 T13177	soq protein - frui
19	55	50.0	1822	2 S63985	collagen alpha 2 c
20	55	50.0	3198	2 A43426	androgen-regulated
21	54.5	49.5	789	2 S28259	collagen alpha 2 f
22	54	49.1	810	2 T10756	collagen alpha 1(I
23	54	49.1	1487	1 CGHU6C	platelet aggregati
24	53.5	48.6	209	2 JX0266	high-sulfur wool m
25	53	48.2	162	2 T47107	keratin high-sulfu
26	53	48.2	172	1 KRSHHA	chordin precursor
27	53	48.2	941	1 A55195	thrombospondin - b
28	52	47.3	84	2 A61594	high-sulfur wool m
29	52	47.3	172	2 147106	high-sulfur wool m

30 52 47.3 1464 1 CGHUIS collagen alpha 1(I
31 51.5 46.8 419 2 S41607 atrolysin A (EC 3.
32 51 46.4 153 2 A27179 collagen alpha 1(I
33 51 46.4 156 1 KRSHHB keratin high-sulfu
34 51 46.4 182 1 KRSHHD keratin high-sulfu
35 51 46.4 182 2 I47105 high-sulfur wool m
36 51 46.4 488 2 A27353 collagen alpha 1(I
37 51 46.4 1466 1 CGHU7L collagen alpha 1(I
38 51 46.4 1486 1 B40333 collagen alpha 1(I
39 51 46.4 1492 2 A40333 collagen alpha 1(I
40 51 46.4 1617 2 T28153 complement C4 - ch
41 49.5 45.0 175 2 S37649 high-sulfur kerati
42 49.5 45.0 177 2 S37650 F1 protein - avian
43 49 44.5 3951 1 VFIBH1 mucin 5AC (clone L
44 48.5 44.1 1042 2 A57534 mucin MUC5B, trach
45 48.5 44.1 1056 2 A53767

RESULT 1
A42587
thrombospondin 2 precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
C:Accession: A42587; A39851
R:Lahterty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during ce
A:Reference number: A42587; MUID:92147683
A:Accession: A42587
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1172 <LAH>
A:Cross-references: GB:L07803; GB:M87275; NID:g340421; PIDN:AAA53064.1; PID:g567241
A:Note: sequence extracted from NCBI backbone (NCBIP:81502)
R:Porststein, P.; O'Rourke, K.; Wikstrom, K.; Wolf, F.W.; Katz, R.; Li, P.; Dixit, V.M.
J. Biol. Chem. 266, 12821-12824, 1991
A:Title: A second, expressed thrombospondin gene (Thbs2) exists in the mouse genome.
A:Reference number: A39851; MUID:91302287
A:Accession: A39851
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-873 <BOR>
A:Cross-references: GB:M64866; NID:g201994; PIDN:AAA40432.1; PID:g201995
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;
C:Keywords: calcium binding; glycoprotein
F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>
F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>
F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>
F:553-588/Domain: EGF homology <EGF1>
F:652-691/Domain: EGF homology <EGF>

Query Match 100.0%; Score 110; DB 2; Length 1172;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PATCANP5FVEGECPCSC 18
|||||
Db 357 PATCANP5FVEGECPCSC 374

RESULT 2
TSHUP2
thrombospondin 2 precursor - human
C:Species: Homo sapiens (man)
C>Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999
C:Accession: A47379; A42173
R:LaBell, T.L.; Byers, P.H.
Genomics 17, 225-229, 1993

ALIGNMENTS

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:24:52 ; Search time 49.78 Seconds
(without alignments)
33.665 Million cell updates/sec

Title: US-09-536-087-8

Perfect score: 128

Sequence: 1 FAENETWVVDSTCTCKKFT 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	128	100.0	1172	1 TSHUP2	thrombospondin 2 p
2	128	100.0	1172	2 A42587	thrombospondin 2 p
3	86	67.2	1178	1 A39804	thrombospondin pre
4	68	53.1	1170	1 TSHUP1	thrombospondin 1 p
5	68	53.1	1170	2 A40558	thrombospondin 1 p
6	64	50.0	1822	2 S63985	collagen alpha 2 c
7	63	49.2	1376	2 S63986	collagen alpha 5 c
8	60	46.9	884	2 T18649	hypothetical prote
9	57	44.5	114	2 A34567	beta-microseminopr
10	57	44.5	3198	2 A43426	collagen alpha 2 f
11	54.5	42.6	5376	2 T42215	zonadhesin - mouse
12	54	42.2	114	2 A54663	seminal plasma pro
13	53	41.4	1535	2 S46224	peroxidasin - frui
14	52	40.6	3020	2 A43932	mucin 2 precursor,
15	50.5	39.5	810	2 T10756	Nel-homolog protei
16	49	38.3	837	2 A42112	mucin-like peptide
17	49	38.3	1291	2 T21694	hypothetical prote
18	48.5	37.9	395	2 A47701	aspartic proteinas
19	47.5	37.1	835	2 JP0076	nel protein - chic
20	47	36.7	269	2 S36166	paired box transcr
21	47	36.7	364	2 T07880	hypothetical prote
22	47	36.7	2813	1 VWHU	von Willebrand fac
23	47	36.7	3026	2 T28431	variant surface pr
24	47	36.7	4845	2 T31067	BIR repeat contain
25	46	35.9	86	2 A25700	trans-activating t
26	46	35.9	101	2 T09446	tat protein - huma
27	46	35.9	966	1 P1BVB8	RNA la protein - b
28	46	35.9	1042	2 A57534	mucin 5AC (clone L
29	46	35.9	1056	2 A53767	mucin MUC5B, trach

30	46	35.9	2910	2 T42214	otocellin - mouse
31	45.5	35.5	157	2 P05663	von Willebrand fac
32	45.5	35.5	590	2 E85057	probable transposo
33	45	35.2	373	2 T45043	hypothetical prote
34	45	35.2	392	2 T32524	hypothetical prote
35	45	35.2	489	1 S47076	3-oxoacyl-(acyl-ca
36	45	35.2	977	2 T16232	hypothetical prote
37	45	35.2	1034	2 JC5598	mucin - rat
38	45	35.2	1038	2 T13177	sugn protein - frul
39	45	35.2	1131	2 T14517	hypothetical prote
40	45	35.2	1188	2 C71231	hypothetical prote
41	44.5	34.8	199	2 T28981	hypothetical prote
42	44.5	34.8	328	2 C72370	tryptophan--tRNA 1
43	44.5	34.8	429	2 T21113	hypothetical prote
44	44.5	34.8	531	2 T20763	hypothetical prote
45	44.5	34.8	629	2 T48287	hypothetical prote

ALIGNMENTS

RESULT 1

TSHUP2

thrombospondin 2 precursor - human

C:Species: Homo sapiens (man)

C:Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999

C:Accession: A47379; A42173

R:LaBell, T.L.; Byers, P.H.

Genomics 17, 225-229, 1993

A>Title: Sequence and characterization of the complete human thrombospondin 2 cDNA: p

A:Reference number: A47379; MUID:94010892

A:Accession: A47379

A:Molecule type: mRNA

A:Residues: 1-1172 <LAB>

A:Cross-references: GB:L12350; NID:g307505; PIDN:AAA03703.1; PID:g307506

R:LaBell, T.L.; Milewicz, D.J.; Distech, C.M.; Byers, P.H.

Genomics 12, 421-429, 1992

A>Title: Thrombospondin II: partial cDNA sequence, chromosome location, and expressio

A:Reference number: A42173; MUID:92217961

A:Accession: A42173

A:Molecule type: mRNA

A:Residues: 560-1172 <LA2>

A:Cross-references: GB:M81339

A:Experimental source: fibroblast

A>Note: sequence extracted from NCBI backbone (NCBIN:95091, NCBIP:95096)

C:Genetics:

A:Gene: GDB:THBS2; TSP2

A:Cross-references: GDB:128789; OMIM:188061

A:Map position: 6q27-6q27

C:Complex: homotrimer, disulfide linked

C:Function:

A:Description: participates in cell migration and adhesion, and in platelet aggregati

C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;

C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tri

F:1-18/Domain: signal sequence status predicted <SIG>

F:19-1172/Product: thrombospondin 2 #status predicted <MAT>

F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>

F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>

F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>

F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>

F:553-588/Domain: EGF homology <EGF1>

F:652-691/Domain: EGF homology <EGF>

F:928-930/Region: cell attachment (R-G-D) motif

F:151,316,330,457,584,710,1069/Binding site: carbohydrate (Asn) (covalent) #status pr

F:167-226/Disulfide bonds: #status predicted

F:266,270/Disulfide bonds: Interchain #status predicted

F:612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 100.0%; Score 128; DB 1; Length 1172;
Best Local Similarity 100.0%; Pred. No. 3.3e-09;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

THIS PAGE BLANK (USPTO)

QY 1 FAENETWVVDSCCTCTCKKFKT 22
Db 327 FAENETWVVDSCCTCTCKKFKT 348

RESULT 2
A42587
thrombospondin 2 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
C:Accession: A42587; A39851
R:Laberty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during cell
A:Reference number: A42587; MUID:92147683
A:Accession: A42587
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1172 <LAH>
A:Cross-references: GB:L07803; GB:M87275; NID:g340421; PIDN:AAA53064.1; PID:g567241
A:Note: sequence extracted from NCBI backbone (NCBIP:81502)
R:Bornstein, P.; O'Rourke, K.; Wikstrom, K.; Wolf, F.W.; Katz, R.; Li, P.; Dixit, V.M.
J. Biol. Chem. 266, 12821-12824, 1991
A:Title: A second, expressed thrombospondin gene (Thbs2) exists in the mouse genome.
A:Reference number: A39851; MUID:91302287
A:Accession: A39851
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-873 <BOR>
A:Cross-references: GB:M64866; NID:g201994; PIDN:AAA0432.1; PID:g201995
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; v
C:Keywords: calcium binding; glycoprotein
F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>
F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>
F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>
F:553-588/Domain: EGF homology <EGF1>
F:652-691/Domain: EGF homology <EGF>

Query Match 100.0%; Score 128; DB 2; Length 1172;
Best Local Similarity 100.0%; Pred. No. 3.3e-09; Mismatches 0; Indels 0; Gaps 0;
Matches 22; Conservative 0;

QY 1 FAENETWVVDSCCTCTCKKFKT 22
Db 327 FAENETWVVDSCCTCTCKKFKT 348

RESULT 3
A39804
thrombospondin precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A39804
R:Lawler, J.; Duquette, M.; Ferro, P.
J. Biol. Chem. 266, 8039-8043, 1991
A:Title: Cloning and sequencing of chicken thrombospondin.
A:Reference number: A39804; MUID:91217026
A:Accession: A39804
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1178 <LAH>
A:Cross-references: GB:M60853; NID:g212763; PIDN:AAA51437.1; PID:g212764
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; v
F:325-383/Domain: von Willebrand factor type C repeat homology <VWC>
F:386-437/Domain: thrombospondin type 1 repeat homology <THR1>
F:442-498/Domain: thrombospondin type 1 repeat homology <THR2>
F:499-555/Domain: thrombospondin type 1 repeat homology <THR3>
F:658-697/Domain: EGF homology <EGF>

Query Match 67.2%; Score 86; DB 1; Length 1178;

Thu Ja-

Best Local Similarity 61.9%; Pred. No. 0.00099;
Matches 13; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
QY 1 FAENETWVVDSCCTCTCKKFKT 21
Db 333 FADSEWVDSCCTCTCQDSK 353

RESULT 4
TSHUP1
thrombospondin 1 precursor - human
C:Species: Homo sapiens (man)
C:Date: 23-Aug-1987 #sequence_revision 03-Aug-1995 #text_change 17-Nov-2000
C:Accession: A26155; A34274; A30140; A25812; A05172; A42927
R:Lawler, J.; Hynes, R.O.
J. Cell Biol. 103, 1635-1648, 1986
A:Title: The structure of human thrombospondin, an adhesive glycoprotein with multipl
A:Reference number: A26155; MUID:87057617
A:Accession: A26155
A:Molecule type: mRNA
A:Residues: 1-1170 <LAH>
A:Cross-references: GB:X04665; NID:g37137; PIDN:CAA28370.1; PID:g37138
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Laberty, C.D.; Gierman, T.M.; Dixit, V.M.
J. Biol. Chem. 264, 11222-11227, 1989
A:Title: Characterization of the promoter region of the human thrombospondin gene. DN
A:Reference number: A34274; MUID:89291870
A:Accession: A34274
A:Molecule type: DNA
A:Residues: 1-166 <LAH>
A:Cross-references: GB:J04835
R:Hennessy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.M.; Rotwel
J. Cell Biol. 108, 729-736, 1989
A:Title: Complete thrombospondin mRNA sequence includes potential regulatory sites in
A:Reference number: A30140; MUID:89139590
A:Accession: A30140
A:Molecule type: mRNA
A:Residues: 1-83, 'A', 85-522, 'A', 524-1170 <HEN>
A:Cross-references: EMBL:X14787; NID:g37464; PIDN:CAA32889.1; PID:g37465
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Kobayashi, S.; Eden-McCutchan, F.; Framson, P.; Bornstein, P.
Biochemistry 25, 8418-8425, 1986
A:Title: Partial amino acid sequence of human thrombospondin as determined by analysi
A:Reference number: A25812; MUID:87157592
A:Accession: A25812
A:Molecule type: mRNA
A:Residues: 1-83, 'A', 85-397 <ROB>
A:Cross-references: GB:M25631; NID:g538353; PIDN:AAA36741.1; PID:g538354
R:Dixit, V.M.; Hennessy, S.W.; Grant, G.A.; Rotwein, P.; Frazier, W.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 5449-5453, 1986
A:Reference number: A05172; MUID:86287276
A:Accession: A05172
A:Molecule type: mRNA
A:Residues: 1-83, 'A', 85-374, 'RC', <DIX>
A:Cross-references: GB:M14326; NID:g340005; PIDN:AAA61237.1; PID:g553801
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Sun, X.; Skorsteingard, K.; Mosher, D.F.
J. Cell Biol. 118, 693-701, 1992
A:Title: Disulfides modulate RGD-inhibitable cell adhesive activity of thrombospondin
A:Reference number: A42927; MUID:92348511
A:Accession: A42927
A:Molecule type: protein
A:Residues: 987-1003 <SUN>
A:Note: Cys-992 is shown to have a free sulfhydryl
C:Genetics:
A:Gene: GDB:THBS1; TSP1; TSP
A:Cross-references: GDB:I120438; OMIM:188060
A:Map position: 15q15-15q15
A:Introns: 23/1
A:Note: the list of introns may be incomplete
C:Complex: homotrimer, disulfide linked
A:Function:
A:Description: participates in cell migration and adhesion, and in platelet aggregati

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:24:54 ; Search time 49.78 Seconds
(without alignments)
22.953 Million cell updates/sec

Title: US-09-536-087-9
Perfect score: 82
Sequence: 1 ELIGGPKTRMSAC 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	1172	1 TSHUP2	thrombospondin 2 p
2	72	87.8	1172	2 A42587	thrombospondin 2 p
3	44	53.7	614	1 W1WL41	El protein - human
4	42	51.2	432	2 S38893	uracil transport p
5	40.5	49.4	958	2 T26258	hypothetical prote
6	40	48.8	514	2 T48987	hypothetical prote
7	40	48.8	673	2 A45456	NADH dehydrogenase
8	39	47.6	81	2 B70966	hypothetical prote
9	39	47.6	439	1 DERSHA	homoserine dehydro
10	39	47.6	604	2 S36493	El protein - human
11	39	47.6	605	2 S36469	El protein - human
12	39	47.6	651	2 A96781	unknown protein F9
13	39	47.6	731	2 T15664	hypothetical prote
14	39	47.6	2437	2 S42612	transmembrane prot
15	38.5	47.0	391	2 T34935	hypothetical prote
16	38	46.3	149	2 I38905	ATP-binding casset
17	38	46.3	207	2 T30242	cystathione synth
18	38	46.3	260	2 C83362	hypothetical prote
19	38	46.3	363	2 T23609	hypothetical prote
20	38	46.3	414	2 D64111	uracil transport p
21	38	46.3	434	2 A57986	membrane-bound ura
22	38	46.3	470	2 D69012	conserved hypothet
23	38	46.3	552	2 E72283	alpha-galactosid
24	38	46.3	556	2 G86319	hypothetical prote
25	38	46.3	1024	2 T16491	hypothetical prote
26	38	46.3	1964	2 T09059	notch4 - mouse
27	37.5	45.7	568	2 E96648	hypothetical prote
28	37	45.1	128	2 C86179	hypothetical prote
29	37	45.1	159	2 A85022	probable actin pol

ALIGNMENTS

RESULT 1

TSHUP2

thrombospondin 2 precursor - human

C:Species: Homo sapiens (man)

C>Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999

C:Accession: A47379; A42173

R:Label: T.L.; Byers, P.H.

Genomics 17, 225-229, 1993

A:Title: Sequence and characterization of the complete human thrombospondin 2 cDNA: P

A:Reference number: A47379; MUID:94010892

A:Accession: A47379

A:Molecule type: mRNA

A:Residues: 1-1172 <LAB>

A:Cross-references: GB:112350; MID:g307505; PID:AAA03703.1; PID:g307506

R:Label: T.L.; Milewicz, D.J.; Distche, C.M.; Byers, P.H.

Genomics 12, 421-429, 1992

A:Title: Thrombospondin II: partial cDNA sequence, chromosome location, and expressio

A:Reference number: A42173; MUID:92217961

A:Accession: A42173

A:Molecule type: mRNA

A:Residues: 560-1172 <LA2>

A:Cross-references: GB:M81339

A:Experimental source: fibroblast

A:Note: sequence extracted from NCBI backbone (NCBIN:95091, NCBI:95096)

C:Genetics:

A:Gene: GDB:THBS2; TSP2

A:Cross-references: GDB:128789; OMIM:188061

A:Map position: 6q27-qq27

C:Complex: homotrimer, disulfide linked

C:Function:

A:Description: participates in cell migration and adhesion, and in platelet aggregati

C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; tri

C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tri

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-1172/Product: von Willebrand factor type C repeat homology <WAT>

F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>

F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>

F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>

F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>

F:553-588/Domain: EGF homology <EGF1>

F:652-691/Domain: EGF homology <EGF2>

F:928-930/Region: cell attachment (R-G-D) motif

F:151-316-330, 457-584, 710, 1069/Binding site: carbohydrate (Asn) (covalent) #status p

F:167-226/Disulfide bonds: #status predicted

F:266,270/Disulfide bonds: interchain #status predicted

F:612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 100.0%; Score 82; DB 1; Length 1172;

Best Local Similarity 100.0%; Pred. No. 4.1e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

THIS PAGE BLANK (USPTO)

QY 1 ELIGPPKTRNMSAC 15
 Db 306 ELIGPPKTRNMSAC 320

RESULT 2
 A42587
 thrombospondin 2 precursor - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
 C:Accession: A42587; A39851
 R:Laherty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
 J. Biol. Chem. 267, 3274-3281, 1992
 A:Title: Characterization of mouse thrombospondin 2 sequence and expression during cell
 A:Reference number: A42587; MUID:92147683
 A:Accession: A42587
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-1172 <LAH>
 A:Cross-references: GB:L07803; GB:M87275; NID:g340421; PIDN:AAA53064.1; PID:g567241
 A:Note: sequence extracted from NCBI backbone (NCBIP:81502)
 R:Bornstein, P.; O'Rourke, K.; Wikstrom, K.; Wolf, F.W.; Katz, R.; Li, P.; Dixit, V.M.
 J. Biol. Chem. 266, 12821-12824, 1991
 A:Title: A second, expressed thrombospondin gene (Thbs2) exists in the mouse genome.
 A:Reference number: A39851; MUID:91302287
 A:Accession: A39851
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-873 <BOR>
 A:Cross-references: GB:IM64866; NID:g201994; PIDN:AAA40432.1; PID:g201995
 C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; v
 F:319-377/Domain: calcium binding; glycoprotein
 F:380-431/Domain: von Willebrand factor type C repeat homology <VWC>
 F:436-492/Domain: thrombospondin type 1 repeat homology <THR1>
 F:493-549/Domain: thrombospondin type 1 repeat homology <THR2>
 F:553-588/Domain: EGF homology <EGF1>
 F:652-691/Domain: EGF homology <EGF>

Query Match 87.8%; Score 72; DB 2; Length 1172;
 Best Local Similarity 93.3%; Pred. No. 0.00023;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ELIGPPKTRNMSAC 15
 Db 306 ELIGPLKTRNMSAC 320

RESULT 3
 W1W141
 E1 protein - human papillomavirus type 41
 C:Species: human papillomavirus type 41
 A:Note: host Homo sapiens (man)
 C>Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jul-1999
 C:Accession: C43550
 R:Hirt, L.; Hirsch-Behnam, A.; De Villiers, E.M.
 Virus Res. 18, 179-190, 1990
 A:Title: Nucleotide sequence of human papillomavirus (HPV) type 41: an unusual HPV type
 A:Reference number: A43550
 A:Accession: C43550
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-614 <HIR>
 A:Cross-references: EMBL:X56147; NID:g60942; PIDN:CAA39614.1; PID:g60945
 C:Superfamily: papillomavirus E1 protein
 C:Keywords: early protein

Query Match 53.7%; Score 44; DB 1; Length 614;
 Best Local Similarity 53.8%; Pred. No. 9.5;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 LIGGPPKTRNMSA 14
 Db 270 MLANPPKTRNMPA 282

RESULT 4
 S38893
 uracil transport protein - Bacillus caldolyticus
 N:Alternate names: uracil permease
 C:Species: Bacillus caldolyticus
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-Sep-1999
 C:Accession: S38893
 R:Chim, S.-Y.; Neuhard, J.
 submitted to the EMBL Data Library, November 1993
 A:Description: The pyrimidine biosynthesis operon of the thermophile Bacillus caldoly
 A:Reference number: S38892
 A:Accession: S38893
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-432 <GHI>
 A:Cross-references: EMBL:X76083; NID:g431229; PIDN:CAA53697.1; PID:g431231
 C:Superfamily: uracil transport protein uraa
 C:Keywords: transmembrane protein

Query Match 51.2%; Score 42; DB 2; Length 432;
 Best Local Similarity 87.5%; Pred. No. 15;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LIGGPPKPT 9
 Db 284 LLGGPPKPT 291

RESULT 5
 T26258
 hypothetical protein W07A8.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T26258
 R:Basham, V.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z20184
 A:Accession: T26258
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-958 <WIL>
 A:Cross-references: EMBL:Z82075; PIDN:CAB04929.1; GSPDB:GN00023; CESP:W07A8.3
 A:Experimental source: clone W07A8
 C:Genetics:
 A:Gene: CESP:W07A8.3
 A:Map position: 5
 A:Introns: 17/2; 35/3; 78/3; 133/2; 167/3; 188/1; 244/3; 359/3; 403/2; 436/1; 499/3;

Query Match 49.4%; Score 40.5; DB 2; Length 958;
 Best Local Similarity 40.0%; Pred. No. 62;
 Matches 10; Conservative 2; Mismatches 2; Indels 11; Gaps 1;

QY 1 ELIGG-----PPKTRNMSA 14
 Db 700 DLGGGFPMPKPTTSTPTNPNPSA 724

RESULT 6
 T48987
 hypothetical protein F25L23.50 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Dec-2000
 C:Accession: T48987
 R:D'Angelio, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.
 submitted to the Protein Sequence Database, May 2000
 A:Reference number: Z25012

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:24:55 ; Search time 49.78 Seconds
(without alignments)
10.712 Million cell updates/sec

Title: US-09-536-087-10

Perfect score: 53

Sequence: 1 WSPFAEW 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	1172	1 TSHUP2	thrombospondin 2 p
2	53	100.0	1172	2 A42587	thrombospondin 2 p
3	50	94.3	802	2 T21315	hypothetical prote
4	50	94.3	1170	1 TSHUP1	thrombospondin 1 p
5	50	94.3	1170	2 A40558	thrombospondin 1 p
6	48	90.6	788	2 T25061	hypothetical prote
7	47	88.7	1178	1 A39804	thrombospondin pre
8	44	83.0	1184	2 T09484	cartilage intermed
9	43	81.1	152	2 S31002	gene 57 protein -
10	43	81.1	152	2 G72806	gp57 protein - Myc
11	42	79.2	213	2 T44588	probable transcrip
12	42	79.2	654	2 T29247	hypothetical prote
13	41	77.4	421	2 T03421	trab protein - Agr
14	41	77.4	574	2 B29677	complement C9 prec
15	41	77.4	708	2 G96518	protein T2E6.8 [im
16	41	77.4	736	2 T19366	hypothetical prote
17	41	77.4	957	2 T15976	hypothetical prote
18	41	77.4	1074	2 JC5928	semaphorin F precu
19	41	77.4	1444	2 T18856	angiogenesis inhib
20	40	75.5	169	2 T75595	hypothetical prote
21	40	75.5	387	2 T02780	probable conjugal
22	40	75.5	702	2 B85091	isomylase-like pr
23	40	75.5	990	1 G46335	env polyprotein pr
24	39	73.6	142	2 C75271	hypothetical prote
25	39	73.6	204	2 T15295	hypothetical prote
26	39	73.6	320	2 F64909	hypothetical prote
27	39	73.6	335	2 T39425	hypothetical prote
28	39	73.6	383	2 T30519	dihydroorotase (EC
29	39	73.6	515	2 S67290	probable membrane

complement C9 prec
mucin JER57 - huma
F-spondin precursor
F-spondin rat
hypothetical prote
unc-5 protein - Ca
gene ADAMTS-1 prot
lysozyme (EC 3.2.1
hypothetical prote
conserved hypotet
glyceroldehyde-3-p
N-formyl peptide r
ferredoxin--NADP+
hypothetical prote
probable isoamylas

30 39 73.6 559 1 C9HU
31 39 73.6 610 2 S35049
32 39 73.6 803 2 A47723
33 39 73.6 807 2 A38152
34 39 73.6 837 2 T00355
35 39 73.6 919 2 T32541
36 39 73.6 947 1 B44294
37 39 73.6 951 2 T00017
38 38 71.7 139 2 S41579
39 38 71.7 254 2 T15952
40 38 71.7 315 2 T44480
41 38 71.7 337 2 S54141
42 38 71.7 352 2 A46520
43 38 71.7 498 1 JT0751
44 38 71.7 651 2 T19477
45 38 71.7 789 2 T01321

ALIGNMENTS

RESULT 1

TSHUP2

thrombospondin 2 precursor - human

C:Species: Homo sapiens (man)

C>Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999

C:Accession: A47379; A42173

R:LaBell, T.L.; Byers, P.H.

Genomics 17, 225-229, 1993

A:Title: Sequence and characterization of the complete human thrombospondin 2 cDNA: p

A:Reference number: A47379; MUID:94010892

A:Accession: A47379

A:Molecule type: mRNA

A:Residues: 1-1172 <LAB>

A:Cross-references: GB:L12350; NID:g307505; PIDN:AAA03703.1; PID:g307506

R:LaBell, T.L.; Milewicz, D.J.; Distche, C.M.; Byers, P.H.

Genomics 12, 421-429, 1992

A:Title: Thrombospondin II: partial cDNA sequence, chromosome location, and expressio

A:Reference number: A42173; MUID:92217961

A:Accession: A42173

A:Molecule type: mRNA

A:Residues: 560-1172 <LA2>

A:Cross-references: GB:M81339

A:Experimental source: fibroblast

A:Note: sequence extracted from NCBI backbone (NCBIN:95091, NCBI:P:95096)

C:Genetics:

A:Gene: GDB:THBS2; TSP2

A:Cross-references: GDB:128789; OMIM:188061

A:Map position: 6q27-6q27

C:Complex: homotrimer, disulfide linked

C:Function:

A:Description: participates in cell migration and adhesion, and in platelet aggregati

C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;

C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tri

F:1-18/Domain: signal sequence #status predicted <SIG>

F:319-377/Domain: von Willebrand factor type C repeat homology <WVC>

F:380-431/Domain: thrombospondin type 1 repeat homology <THRI>

F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>

F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>

F:553-588/Domain: EGF homology <EGF1>

F:652-691/Domain: EGF homology <EGF>

F:928-930/Region: cell attachment (R-G-D) motif

F:151,316,330,457,584,710,1069/Binding site: carbohydrate (Asn) (covalent) #status pr

F:167-226/Disulfide bonds: #status predicted

F:266,270/Disulfide bonds: interchain #status predicted

F:612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 100.0%; Score 53; DB 1; Length 1172;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

THIS PAGE BLANK (uspto)

QY 1 WSPWAEW 7
|||||
Db 384 WSPWAEW 390

RESULT 2
A42587
thrombospondin 2 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
C:Accession: A42587; A39851
R:Raherty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during cell
A:Reference number: A42587; MUID:92147683
A:Accession: A42587
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1172 <LAH>
A:Cross-references: GB:L07803; GB:M87275; NID:G340421; PIDN:AAA53064.1; PID:G567241
A:Note: sequence extracted from NCBI backbone (NCBIP:81502)
R:Bornstein, P.; O'Rourke, K.; Wikstrom, K.; Wolf, F.W.; Katz, R.; Li, P.; Dixit, V.M.
J. Biol. Chem. 266, 12821-12824, 1991
A:Title: A second, expressed thrombospondin gene (Thbs2) exists in the mouse genome.
A:Reference number: A39851; MUID:91302287
A:Accession: A39851
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-873 <BOR>
A:Cross-references: GB:M64866; NID:G201994; PIDN:AAA40432.1; PID:G201995
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; von
F:319-377/Domain: von Willebrand factor type C repeat homology <WYC>
F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>
F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>
F:553-588/Domain: EGF homology <EGF1>
F:652-691/Domain: EGF homology <EGF>

Query Match 100.08; Score 53; DB 2; Length 1172;
Best Local Similarity 100.08; Pred. No. 1.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
|||||
Db 384 WSPWAEW 390

RESULT 3
T21315
hypothetical protein F23H12.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21315
R:Kershaw, J.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19405
A:Accession: T21315
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-802 <WIL>
A:Cross-references: EMBL:Z74472; PIDN:CAA98943.1; GSPDB:GNO0023; CESP:F23H12.5
A:Experimental source: clone F23H12
C:Genetics:
A:Gene: CESP:F23H12.5
A:Map position: 5
A:Introns: 24/1; 64/3; 109/1; 133/1; 522/3; 548/3; 663/3; 704/1; 740/1

Query Match 94.38; Score 50; DB 2; Length 802;
Best Local Similarity 85.78; Pred. No. 3.4;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
|||||
Db 706 WSPWSEW 712

RESULT 4
TSHUP1
thrombospondin 1 precursor - human
C:Species: Homo sapiens (man)
C:Date: 23-Aug-1987 #sequence_revision 03-Aug-1995 #text_change 17-Nov-2000
C:Accession: A26155; A32474; A30140; A25812; A05172; A42927
R:Lawler, J.; Hynes, R.O.
J. Cell Biol. 103, 1635-1648, 1986
A:Title: The structure of human thrombospondin, an adhesive glycoprotein with multiple
A:Reference number: A26155; MUID:87057617
A:Accession: A26155
A:Molecule type: mRNA
A:Residues: 1-1170 <LAW>
A:Cross-references: GB:X04665; NID:G37137; PIDN:CAA28370.1; PID:G37138
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Lawler, C.D.; Gierman, T.M.; Dixit, V.M.
J. Biol. Chem. 264, 11222-11227, 1989
A:Title: Characterization of the promoter region of the human thrombospondin gene. DN
A:Reference number: A34274; MUID:89291870
A:Accession: A34274
A:Molecule type: DNA
A:Residues: 1-166 <LAH>
A:Cross-references: GB:J04835
R:Hennessy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.M.; Rotwei
J. Cell Biol. 108, 729-736, 1989
A:Title: Complete thrombospondin mRNA sequence includes potential regulatory sites in
A:Reference number: A30140; MUID:89139590
A:Accession: A30140
A:Molecule type: mRNA
A:Residues: 1-83, 'A', '85-522, 'A', '524-1170 <HEN>
A:Cross-references: EMBL:X14787; NID:G37464; PIDN:CAA32889.1; PID:G37465
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Kobayashi, S.; Eden-McCutchan, F.; Franson, P.; Bornstein, P.
Biochemistry 25, 8418-8425, 1986
A:Title: Partial amino acid sequence of human thrombospondin as determined by analysis
A:Reference number: A25812; MUID:87157592
A:Accession: A25812
A:Molecule type: mRNA
A:Residues: 1-83, 'A', '85-397 <KOB>
A:Cross-references: GB:M25631; NID:G538353; PIDN:AAA36741.1; PID:G538354
R:Dixit, V.M.; Hennessy, S.W.; Grant, G.A.; Rotwein, P.; Frazier, W.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 5449-5453, 1986
A:Reference number: A05172; MUID:86287276
A:Accession: A05172
A:Molecule type: mRNA
A:Residues: 1-83, 'A', '85-374, 'RC' <DIX>
A:Cross-references: GB:M14326; NID:G340005; PIDN:AAA61237.1; PID:G553801
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Sun, X.; Skorstengaard, K.; Mosher, D.F.
J. Cell Biol. 118, 693-701, 1992
A:Title: Disulfides modulate RGD-inhibitable cell adhesive activity of thrombospondin
A:Reference number: A42927; MUID:92348511
A:Accession: A42927
A:Molecule type: protein
A:Residues: 987-1003 <SUN>
A:Note: Cys-992 is shown to have a free sulfhydryl
C:Genetics:
A:Gene: GDB:THBS1; TSP1; TSP
A:Cross-references: GDB:120438; OMIM:188060
A:Map position: 15q15-15q15
A:Introns: 23/1
A:Note: the list of introns may be incomplete
C:Complex: homotrimer, disulfide linked
C:Function:
A:Description: participates in cell migration and adhesion, and in platelet aggregati
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;
C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tri

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:20:35 ; Search time 49.78 seconds
(without alignments)
24.484 Million cell updates/sec

Title: US-09-536-087-6

Perfect score: 84

Sequence: 1 RESHFRGLQNVHLVF 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84	100.0	1172	1 TSHUP2	thrombospondin 2 p
2	84	100.0	1172	2 A42587	thrombospondin 2 p
3	79	94.0	1178	1 A39804	thrombospondin pre
4	44	52.4	229	2 S57957	thrombospondin 1 -
5	44	52.4	1170	1 TSHUP1	thrombospondin 1 p
6	44	52.4	1170	2 A40558	thrombospondin 1 p
7	43	51.2	350	2 T46957	hypothetical prote
8	43	51.2	1383	2 T06091	hypothetical prote
9	42	50.0	361	2 T29571	hypothetical prote
10	41	48.8	177	2 E84245	inosine-5'-monopho
11	41	48.8	452	2 T28883	hypothetical prote
12	41	48.8	693	2 C86364	hypothetical prote
13	40	47.6	397	2 B70505	hypothetical prote
14	40	47.6	501	2 T52170	cytochrome P450 mo
15	40	47.6	614	2 T29902	hypothetical prote
16	40	47.6	752	2 T48574	hypothetical prote
17	39	46.4	201	2 E72321	hypothetical prote
18	39	46.4	410	1 DEECPG	phosphoglycerate d
19	39	46.4	410	2 D85947	hypothetical prote
20	39	46.4	865	2 T40288	hypothetical prote
21	39	46.4	1245	2 H83574	conserved hypotnet
22	39	46.4	1286	2 T71413	hypothetical prote
23	38.5	45.8	136	2 T33423	hypothetical prote
24	38.5	45.8	4466	1 S17231	dyein beta heavy
25	38.5	45.8	4466	1 S17653	dyein beta heavy
26	38	45.2	249	2 H72200	conserved hypotnet
27	38	45.2	351	2 H72328	cell division prot
28	38	45.2	622	2 S56214	probable membrane
29	38	45.2	750	1 JDVLVH	DNA-directed DNA p

30	38	45.2	830	2 B44439	protein kinase (EC
31	38	45.2	832	2 S71785	DNA-directed DNA p
32	38	45.2	856	2 A44439	protein kinase (EC
33	38	45.2	1247	2 T18671	hypothetical prote
34	38	45.2	1417	2 A57570	Bloom's syndrome r
35	38	45.2	2082	2 T37056	probable multi-dom
36	38	45.2	4836	2 T14346	hcr2 protein - mo
37	37.5	44.6	641	2 S54064	probable membrane
38	37	44.0	82	2 F96625	hypothetical prote
39	37	44.0	117	2 T13101	minor tail protein
40	37	44.0	208	2 C81800	phosphoribosylglyc
41	37	44.0	212	1 XVECGF	phosphoribosylglyc
42	37	44.0	212	2 F85893	phosphoribosylglyc
43	37	44.0	243	2 T25746	hypothetical prote
44	37	44.0	252	2 H75123	thiamin biosynthet
45	37	44.0	255	2 G71007	thiamin biosynthes

ALIGNMENTS

RESULT 1

TSHUP2

thrombospondin 2 precursor - human

C:Species: Homo sapiens (man)

C:Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999

C:Accession: A47379; A42173

R:Label: T.L.; Byers, P.H.

Genomics 17, 225-229, 1993

A:Title: Sequence and characterization of the complete human thrombospondin 2 cDNA: p

A:Reference number: A47379; MUID:94010892

A:Accession: A47379

A:Molecule type: mRNA

A:Residues: 1-1172 <LAB>

A:Cross-references: GB:L12350; NID:g307505; PIDN:AAA03703.1; PID:g307506

R:Label: T.L.; Milewicz, D.J.; Distche, C.M.; Byers, P.H.

Genomics 12, 421-429, 1992

A:Title: Thrombospondin II: partial cDNA sequence, chromosome location, and expressio

A:Reference number: A42173; MUID:92217961

A:Accession: A42173

A:Molecule type: mRNA

A:Residues: 560-1172 <LA2>

A:Cross-references: GB:M81339

A:Experimental source: fibroblast

A:Note: sequence extracted from NCBI backbone (NCBIN:95091, NCBI:95096)

C:Genetics: TSP2

A:Gene: GDB:THBS2; TSP2

A:Cross-references: GDB:128789; OMIM:188061

A:Map position: 6q27-6q27

C:Complex: homotrimer, disulfide linked

C:Function:

A:Description: participates in cell migration and adhesion, and in platelet aggregati

C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;

C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tri

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-1172/Product: thrombospondin 2 #status predicted <MAT>

F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>

F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>

F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>

F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>

F:553-588/Domain: EGF homology <EGF1>

F:652-691/Domain: EGF homology <EGF>

F:928-930/Region: cell attachment (R-G-D) motif

F:151,316,330,457,584,710,1069/Binding site: carboxydrate (Asn) (covalent) #status pr

F:167-226/Disulfide bonds: #status predicted

F:266,270/Disulfide bonds: interchain #status predicted

F:612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 100.0%; Score 84; DB 1; Length 1172;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESHFRGLQNVLVLF 16
 |||
 Db 198 RESHFRGLQNVLVLF 213

RESULT 2

A42587
 thrombospondin 2 precursor - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
 C:Accession: A42587; A39851
 R:Lawler, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
 J. Biol. Chem. 267, 3274-3281, 1992
 A>Title: Characterization of mouse thrombospondin 2 sequence and expression during cell
 A:Reference number: A42587; MUID:92147683
 A:Accession: A42587
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-1172 <LAH>
 A:Cross-references: GB:M87275; NID:g340421; PIDN:AAA53064.1; PID:g567241
 A:Note: sequence extracted from NCBI backbone (NCBIP:81502)
 R:Bornstein, P.; O'Rourke, K.; Wikstrom, K.; Wolf, F.W.; Katz, R.; Li, P.; Dixit, V.M.
 J. Biol. Chem. 266, 12821-12824, 1991
 A>Title: A second, expressed thrombospondin gene (Thbs2) exists in the mouse genome.
 A:Reference number: A39851; MUID:91302287
 A:Accession: A39851
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-873 <BOR>
 A:Cross-references: GB:M64866; NID:g201994; PIDN:AAA0432.1; PID:g201995
 C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; vd
 C:Keywords: calcium binding; glycoprotein
 F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>
 F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
 F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>
 F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>
 F:553-588/Domain: EGF homology <EGF1>
 F:652-691/Domain: EGF homology <EGF>

Query Match 100.0%; Score 84; DB 2; Length 1172;

Best Local Similarity 100.0%; Pred. No. 1.2e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESHFRGLQNVLVLF 16
 |||
 Db 198 RESHFRGLQNVLVLF 213

RESULT 3

A39804
 thrombospondin precursor - chicken
 C:Species: Gallus gallus (chicken)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A39804
 R:Lawler, J.; Duquette, M.; Ferro, P.
 J. Biol. Chem. 266, 8039-8043, 1991
 A>Title: Cloning and sequencing of chicken thrombospondin.
 A:Reference number: A39804; MUID:91217026
 A:Accession: A39804
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1178 <LAH>
 A:Cross-references: GB:M60853; NID:g212763; PIDN:AAA51437.1; PID:g212764
 C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; vd
 F:325-383/Domain: von Willebrand factor type C repeat homology <VWC>
 F:386-437/Domain: thrombospondin type 1 repeat homology <THR1>
 F:442-498/Domain: thrombospondin type 1 repeat homology <THR2>
 F:499-555/Domain: thrombospondin type 1 repeat homology <THR3>
 F:658-697/Domain: EGF homology <EGF>

Query Match 94.0%; Score 79; DB 1; Length 1178;

Best Local Similarity 81.2%; Pred. No. 9.6e-06;
 Matches 13; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RESHFRGLQNVLVLF 16
 |||
 Db 204 RENHFRGLQNHLIF 219

RESULT 4

S57957
 thrombospondin 1 - bovine (fragment)
 C:Species: Bos primigenius taurus (cattle)
 C>Date: 13-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 20-Aug-1999
 C:Accession: S57957
 R:Lafuillade, B.; Pellerin, S.; Keramidas, M.; Chambaz, E.M.; Feige, J.J.
 submitted to the EMBL Data Library, July 1995
 A>Description: Opposite regulation of thrombospondin-1 and CISP/thrombospondin-2 expr
 A:Reference number: S57955
 A:Accession: S57957
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-229 <LAH>
 A:Cross-references: EMBL:X89511; NID:g899228; PIDN:CAA61682.1; PID:g899229
 C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;

Query Match 52.4%; Score 44; DB 2; Length 229;

Best Local Similarity 61.5%; Pred. No. 2.6;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 HFRGLQNVLVLF 16
 |||
 Db 189 NFGVLQNVRVFF 201

RESULT 5

TSHUP1
 thrombospondin 1 precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 23-Aug-1987 #sequence_revision 03-Aug-1995 #text_change 17-Nov-2000
 C:Accession: A26155; A34274; A30140; A25812; A05172; A2927
 R:Lawler, J.; Hynes, R.O.
 J. Cell Biol. 103, 1635-1648, 1986
 A>Title: The structure of human thrombospondin, an adhesive glycoprotein with multipl
 A:Reference number: A26155; MUID:87057617
 A:Accession: A26155
 A:Molecule type: mRNA
 A:Residues: 1-1170 <LAH>
 A:Cross-references: GB:X04665; NID:g37137; PIDN:CAA38370.1; PID:g37138
 A:Note: parts of this sequence, including the amino end of the mature protein, were d
 R:Lahterty, C.D.; Gierman, T.M.; Dixit, V.M.
 J. Biol. Chem. 264, 11222-11227, 1989
 A>Title: Characterization of the promoter region of the human thrombospondin gene. DN
 A:Reference number: A34274; MUID:89291870
 A:Accession: A34274
 A:Molecule type: DNA
 A:Residues: 1-166 <LAH>
 A:Cross-references: GB:J04835
 R:Hennessy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.M.; Rotwei
 J. Cell Biol. 108, 729-736, 1989
 A>Title: Complete thrombospondin mRNA sequence includes potential regulatory sites in
 A:Reference number: A30140; MUID:89139590
 A:Accession: A30140
 A:Molecule type: mRNA
 A:Residues: 1-83, A' 85-522, A' 524-1170 <HEN>
 A:Cross-references: EMBL:X14787; NID:g37464; PIDN:CAA32889.1; PID:g37465
 A:Note: parts of this sequence, including the amino end of the mature protein, were d
 R:Kobayashi, S.; Eden-McCutchan, F.; Framson, P.; Bornstein, P.
 Biochemistry 25, 8418-8425, 1986
 A>Title: Partial amino acid sequence of human thrombospondin as determined by analysi
 A:Reference number: A25812; MUID:87157592
 A:Accession: A25812
 A:Molecule type: mRNA

A:Residues: 1-83,'A','85-397 <ROB>
A:Cross-references: GB:M25631; NID:g538353; PIDN:AAA36741.1; PID:g538354
R:Dixit, V.M.; Hennessy, S.W.; Grant, G.A.; Rotwein, P.; Frazier, W.A.
Proc. Natl. Acad. Sci. U.S.A. 83: 5449-5453, 1986
A:Reference number: A05172; MUID:86287276
A:Accession: A05172
A:Molecule type: mRNA
A:Residues: 1-83,'A','85-374','RC' <DX>
A:Cross-references: GB:M14326; NID:g340005; PIDN:AAA61237.1; PID:g553801
A:Note: parts of this sequence, including the amino end of the mature protein, were determined by J. Cell Biol. 118, 693-701, 1992
R:Sun, X.; Skorstengaard, K.; Mosher, D.F.
J. Cell Biol. 118, 693-701, 1992
A:Title: Disulfides modulate RGD-inhibitable cell adhesive activity of thrombospondin.
A:Reference number: A42927; MUID:92348511
A:Accession: A42927
A:Molecule type: protein
A:Residues: 987-1003 <SN>
A:Note: Cys-992 is shown to have a free sulphydryl
C:Genetics:
A:Gene: GDB:THBS1; TSP1; TSP
A:Cross-references: GDB:120438; OMIM:188060
A:Map position: 15q15-15q15
A:Introns: 23/1
A:Note: the list of introns may be incomplete
C:Complex: homotrimer, disulfide linked
C:Function:
A:Description: participates in cell migration and adhesion, and in platelet aggregation
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; von Willebrand factor type C repeat homology; calcium binding; glycoprotein; trimer
F:1-16/Domain: signal sequence #status predicted <SIG>
F:19-1170/Product: thrombospondin 1 #status predicted <MAT>
F:317-375/Domain: von Willebrand factor type C repeat homology <THR1>
F:378-429/Domain: thrombospondin type 1 repeat homology <THR2>
F:434-490/Domain: thrombospondin type 1 repeat homology <THR3>
F:491-547/Domain: thrombospondin type 1 repeat homology <EGF1>
F:551-586/Domain: EGF homology <EGF2>
F:650-689/Domain: EGF homology <EGF3>
F:926-928/Region: cell attachment (R-G-D) motif
F:171-232/Disulfide bonds: #status predicted
F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:270,274/Disulfide bonds: interchain #status predicted
F:610/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:1051/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 52.4%; Score 44; DB 1; Length 1170;
Best Local Similarity 61.5%; Pred. No. 16;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 HFRGLQLQNVHLVF 16
:|:|:|:|:|:|
Db 207 NFOGVLQNVRFV 219

RESULT 6
A40558
thrombospondin 1 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 20-Aug-1999
R:Lawler, J.; Duquette, M.; Ferro, P.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.
Genomics 11, 587-600, 1991
A:Title: Characterization of the murine thrombospondin gene.
A:Reference number: A40558; MUID:92128941
A:Accession: A40558
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1170 <LAW>
A:Cross-references: GB:M62449; GB:M62450; GB:M62451; GB:M62452; GB:M62453; GB:M62454; GB:M62455; GB:M62456; GB:M62467; GB:M62468; GB:M62469; GB:M62470; NID:g511867; PIDN:AAA5061
R:Bornstein, P.; Alfi, D.; Devarayalu, S.; Framson, P.; Li, P.
J. Biol. Chem. 265, 16691-16698, 1990
A:Title: Characterization of the mouse thrombospondin gene and evaluation of the role of

A:Reference number: A37905; MUID:90375546
A:Accession: A37905
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-490 <BOR>
A:Cross-references: GB:J05605; GB:J05606; NID:g201991; PIDN:AAA40431.1; PID:g554390
R:Lahterty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during cell differentiation
A:Reference number: A42587; MUID:92147693
A:Accession: B42587
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1152,'P','1154-1170 <LAW>
A:Cross-references: GB:M87276
A:Note: sequence extracted from NCBI backbone (NCBI:81501)
R:Chen, H.; Aeschlimann, D.; Nowien, J.; Mosher, D.F.
FEBS Lett. 387, 36-41, 1996
A:Title: Expression and initial characterization of recombinant mouse thrombospondin
A:Reference number: S68787; MUID:96234006
A:Accession: S68787
A:Molecule type: protein
A:Residues: 19-26,'X','28-37 <CHE>
C:Complex: homotrimer, disulfide linked
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; von Willebrand factor type C repeat homology; glycoprotein; homotrimer
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-1170/Product: thrombospondin 1 #status predicted <MAT>
F:317-375/Domain: von Willebrand factor type C repeat homology <THR1>
F:378-429/Domain: thrombospondin type 1 repeat homology <THR2>
F:434-490/Domain: thrombospondin type 1 repeat homology <THR3>
F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>
F:551-586/Domain: EGF homology <EGF>
F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 52.4%; Score 44; DB 2; Length 1170;
Best Local Similarity 61.5%; Pred. No. 16;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 HFRGLQLQNVHLVF 16
:|:|:|:|:|:|
Db 207 NFOGVLQNVRFV 219

RESULT 7
T46997
hypothetical protein aruF [imported] - Yersinia pestis
C:Species: Yersinia pestis
C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 19-May-2000
C:Accession: T46997
R:Buchrieser, C.; Rusniok, C.; Couve, E.; Prangeul, L.; Billault, A.; Kunst, F.; Carn
submitted to the EMBL Data Library, October 1998
A:Description: DNA sequence of the 102 kbases unstable region of Yersinia pestis.
A:Reference number: Z24348
A:Accession: T46997
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-350 <BUC>
A:Cross-references: EMBL:AL031866; PIDN:CAA21340.1
A:Experimental source: strain 6/69
C:Genetics:
A:Note: aruF
C:Superfamily: Escherichia coli hypothetical protein bl747

Query Match 51.2%; Score 43; DB 2; Length 350;
Best Local Similarity 63.6%; Pred. No. 6.3;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 HFRGLQLQNVHL 14
:|:|:|:|:|:|
Db 305 HFRAILMNTL 315

Q. Now, you said that the defendant was not in the car at the time of the shooting, is that correct?

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:24:52 ; Search time 49.78 Seconds
(without alignments)
27.544 Million cell updates/sec

Title: US-09-536-087-7

Perfect score: 110
Sequence: 1 PATCANPSFVEGCCPSC 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	100.0	1172	2 A42587	thrombospondin 2 p
2	105	95.5	1172	1 TSHUP2	thrombospondin 2 p
3	90	81.8	1178	1 A39804	thrombospondin pre
4	66	60.0	884	2 T18649	hypothetical prote
5	62	56.4	1497	2 I49607	procollagen type V
6	59	53.6	206	2 S18250	collagen alpha 1(I
7	59	53.6	615	2 A05269	collagen alpha 1(I
8	59	53.6	886	2 I50894	collagen alpha 1(I
9	59	53.6	1170	1 TSHUP1	thrombospondin 1 p
10	59	53.6	1170	2 A40558	thrombospondin 1 p
11	59	53.6	1376	2 S63986	collagen alpha 5 c
12	59	53.6	1487	2 B41182	collagen alpha 1(I
13	58.5	53.2	429	2 T21113	hypothetical prote
14	57	51.8	1496	1 CGHU2V	collagen alpha 2(V
15	56	50.9	98	2 I49562	alpha-1 type III c
16	56	50.9	1464	2 S28556	collagen alpha 1(I
17	55.5	50.5	776	2 S28258	androgen-regulated
18	55	50.0	1038	2 T13177	soq protein - frui
19	55	50.0	1822	2 S63985	collagen alpha 2 c
20	55	50.0	3198	2 A34426	collagen alpha 2 f
21	54.5	49.5	789	2 S28259	androgen-regulated
22	54	49.1	810	2 T10756	Nel-homolog protei
23	54	49.1	1487	1 CGHU6C	collagen alpha 1(I
24	53.5	48.6	209	2 JX0266	platelet aggregati
25	53	48.2	162	2 I47107	high-sulfur wool m
26	53	48.2	172	1 KRSHHA	keratin high-sulfu
27	53	48.2	941	1 A55195	chordin precursor
28	52	47.3	84	2 A61594	thrombospondin - b
29	52	47.3	172	2 I47106	high-sulfur wool m

collagen alpha 1(I
atrolysins A (EC 3.
collagen alpha 1(I
keratin high-sulfu
keratin high-sulfu
high-sulfur wool m
collagen alpha 1(I
collagen alpha 1(I
collagen alpha 1(I
collagen alpha 1(I
complement C4 - ch
high-sulfur kerati
high-sulfur kerati
F1 protein - avian
mucin 5AC (clone L
mucin MUC5B, trach

ALIGNMENTS

RESULT 1
A42587
thrombospondin 2 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
C:Accession: A42587; A39851
R:Laherty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during ce
A:Reference number: A42587; MUID:92147683
A:Accession: A42587
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1172 <LAH>
A:Cross-references: GB:L07803; GB:M87275; NID:G340421; PIDN:AAA33064.1; PID:G567241
A:Note: sequence extracted from NCBI backbone (NCBIP:81502)
R:Bornstein, P.; O'Rourke, K.; Wikstrom, K.; Wolf, F.W.; Katz, R.; Li, P.; Dixit, V.M.
J. Biol. Chem. 266, 12821-12824, 1991
A:Title: A second, expressed thrombospondin gene (Thbs2) exists in the mouse genome.
A:Reference number: A39851; MUID:91302287
A:Accession: A39851
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-873 <BOR>
A:Cross-references: GB:M64866; NID:G201994; PIDN:AAA0432.1; PID:G201995
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;
C:Keywords: calcium binding; glycoprotein
F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>
F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>
F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>
F:553-588/Domain: EGF homology <EGF1>
F:652-691/Domain: EGF homology <EGF>

Query Match 100.0%; Score 110; DB 2; Length 1172;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PATCANPSFVEGCCPSC 18
|||||
Db 357 PATCANPSFVEGCCPSC 374

RESULT 2
TSHUP2
thrombospondin 2 precursor - human
C:Species: Homo sapiens (man)
C:Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999
C:Accession: A47379; A42173
R:LaBell, T.L.; Byers, P.H.
Genomics 17, 225-229, 1993

```

Best Local Similarity 72.2%; Pred. NO. 0.00021;
Matches 13; Conservative 2; Mismatches 0; Gaps 0;

QY 1 PATCANPSFVEGECPCSPC 18
   |:::|:::|:::|:::|
DB 363 PVSACDPSFIEGECPCVC 380

RESULT 4
T18649
hypothetical protein B0024.14 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18649
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19001
A:Accession: T18649
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-884 <WIL>
A:Cross-references: EMBL:Z71178; PIDN:CAA94886.1; GSPDB:GN00023; CESP:B0024.14
A:Experimental source: clone B0024
C:Genetics:
A:Gene: CESP:B0024.14
A:Map position: 5
A:Introns: 46/3; 84/1; 212/1; 307/2; 345/2; 394/1; 424/2; 481/1; 596/1; 702/1; 765/3;

Query Match 60.0%; Score 66; DB 2; Length 884;
Best Local Similarity 66.7%; Pred. NO. 0.25;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 CANPSFVEGECPCSPC 18
   | | | | | | | | | |
DB 272 CENPRKVGCCPCVC 286

RESULT 5
I49607
procollagen type V alpha 2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C:Accession: I49607
R:Andrikopoulos, K.; Suzuki, H.R.; Solursh, M.; Ramirez, F.
Dev. Dyn. 195, 113-120, 1992
A:Title: Localization of pro-alpha 2(V) collagen transcripts in the tissues of the
A:Reference number: I49607; MUID:93214071
A:Accession: I49607
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1497 <RES>
A:Cross-references: GB:I02918; NID:g309180; PIDN:AAA37440.1; PID:g309181
C:Genetics:
A:Gene: Col5a-2
A:CSuperfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
F:39-98/Domain: von Willebrand factor type C repeat homology <VMC>
F:1270-1497/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 56.4%; Score 62; DB 2; Length 1497;
Best Local Similarity 66.7%; Pred. NO. 1.3;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 CANPSFVEGECPCSPC 18
   | | | | | | | | | |
DB 81 CANPITPPGECPCVC 95

RESULT 6
S18250
collagen alpha 1(II) chain precursor - chicken (fragment)
C:Species: Gallus gallus (chicken)

```

C:Date: 22-Nov-1993 #sequence_revision 26-Jul-1996 #text_change 13-Aug-1999
 C:Accession: S18250
 R:Nah, H.D.; Upholt, W.B.
 J. Biol. Chem. 266, 23446-23452, 1991
 A:Title: Type II collagen mRNA containing an alternatively spliced exon predominates in
 A:Reference number: S18250; MUID:92078225
 A:Accession: S18250
 A:Molecule type: mRNA
 A:Residues: 1-206 <NAH>
 A:Cross-references: EMBL:M74435; NID:g211635; PIDN:AAA48714.1; PID:g211636
 C:Genetics:
 A:Gene: COL2A1
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-184/Domain: amino-terminal propeptide #status predicted <PRO>
 F:36-95/Domain: von Willebrand factor type C repeat homology <VWC>
 F:185-206/Product: collagen alpha 1(II) chain (fragment) #status predicted <MAT>

Query Match 53.6%; Score 59; DB 2; Length 206;
 Best Local Similarity 50.0%; Pred. No. 0.65;
 Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 PATCANPSFVEGECPC 18
 | | | | | | | | | |
 Db 75 PDCPSPEIPGECPC 92

RESULT 7
 A05269
 collagen alpha 1(III) chain precursor - chicken (fragments)
 C:Species: Gallus gallus (chicken)
 C:Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 16-Jul-1999
 C:Accession: A05269; A38035; A20855
 R:Yamada, Y.; Liaw, G.; Mudryj, M.; Obici, S.; de Crombrughe, B.
 Nature 310, 333-337, 1984
 A:Title: Conservation of the sizes for one but not another class of exons in two chick
 A:Reference number: A05269; MUID:84270696
 A:Accession: A05269
 A:Molecule type: DNA
 A:Residues: 1-329 <YAM1>
 A:Note: the authors translated the codon GAT for residue 548 as Glu
 R:Yamada, Y.; Mudryj, M.; Sullivan, M.; de Crombrughe, B.
 J. Biol. Chem. 258, 2758-2761, 1983
 A:Title: Isolation and characterization of a genomic clone encoding chick alpha type II
 A:Reference number: A20855; MUID:83135706
 A:Accession: A20855
 A:Molecule type: DNA
 A:Residues: 223-233, 'R', '235-236, 'V', 238-239, 330-336, 'V', 338-351 <YAM>
 A:Note: the authors translated the codon CGT for residue 234 as Gly, GTT for residue 237
 C:Comment: Chicken collagen alpha 1(III) chain has about 50 exons. This sequence corres
 , 9(240-275), 6(276-293), 5(294-329), 4(330-423), 3(424-484), 2(485-567), and 1(568-615)
 C:Genetics:
 A:Introns: 68/3; 86/3; 119/3; 134/3; 152/3; 188/3; 221/3; 239/3; 275/3; 293/3; 423/3; 48
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; tandem repeat; trimer; trip
 F:1-68/Region: amino-terminal propeptide nonhelical
 F:2-62/Domain: von Willebrand factor type C repeat homology <VWC>
 F:69-347/Domain: collagen alpha helical chain (fragments) #status predicted <CAH>
 F:348-615/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F:387-615/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
 F:347,348/bisulfide bonds: interchain #status experimental
 F:516/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 53.6%; Score 59; DB 2; Length 615;
 Best Local Similarity 60.0%; Pred. No. 1.5;
 Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 CANPSFVEGECPC 18
 | | | | | | | | | |
 Db 45 CPNPEIPGECPC 59

RESULT 8
 I50694
 collagen alpha 1(III) chain - chicken (fragment)
 C:Species: Gallus gallus (chicken)
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Aug-1999
 C:Accession: I50694
 R:Nah, H.D.; Niu, Z.; Adams, S.L.
 J. Biol. Chem. 269, 16443-16448, 1994
 A:Title: An alternative transcript of the chick type III collagen gene that does not
 A:Reference number: A54041; MUID:94266842
 A:Accession: I50694
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-886 <NAH>
 A:Cross-references: EMBL:U07973; NID:g520454; PIDN:AAA83407.1; PID:g537432
 C:Genetics:
 A:Gene: COL3A1
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
 F:30-90/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 53.6%; Score 59; DB 2; Length 886;
 Best Local Similarity 60.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 CANPSFVEGECPC 18
 | | | | | | | | | |
 Db 73 CPNPEIPGECPC 87

RESULT 9
 TSHUP1
 thrombospondin 1 precursor - human
 C:Species: Homo sapiens (man)
 C:Date: 23-Aug-1987 #sequence_revision 03-Aug-1995 #text_change 17-Nov-2000
 C:Accession: A26155; A34274; A30140; A25812; A05172; A42927
 R:Lawler, J.; Hynes, R.O.
 J. Cell Biol. 103, 1635-1648, 1986
 A:Title: The structure of human thrombospondin, an adhesive glycoprotein with multipl
 A:Reference number: A26155; MUID:87057617
 A:Accession: A26155
 A:Molecule type: mRNA
 A:Residues: 1-1170 <LAW>
 A:Cross-references: GB:X04665; NID:g37137; PIDN:CAA28370.1; PID:g37138
 A:Note: parts of this sequence, including the amino end of the mature protein, were d
 R:Laberty, C.D.; Glerman, T.M.; Dixit, V.M.
 J. Biol. Chem. 264, 11222-11227, 1989
 A:Title: Characterization of the promoter region of the human thrombospondin gene. DN
 A:Reference number: A34274; MUID:89291870
 A:Accession: A34274
 A:Molecule type: DNA
 A:Residues: 1-166 <LAH>
 A:Cross-references: GB:J04835
 R:Hennessy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.M.; Rotwei
 J. Cell Biol. 108, 729-736, 1989
 A:Title: Complete thrombospondin mRNA sequence includes potential regulatory sites in
 A:Reference number: A30140; MUID:89139590
 A:Accession: A30140
 A:Molecule type: mRNA
 A:Residues: 1-83, 'A', '85-522, 'A', 524-1170 <HEN>
 A:Cross-references: EMBL:X14787; NID:g37464; PIDN:CAA32889.1; PID:g37465
 A:Note: parts of this sequence, including the amino end of the mature protein, were d
 R:Kobayashi, S.; Eden-McCutchan, F.; Framson, P.; Bornstein, P.
 Biochemistry 25, 8418-8425, 1986

M62465; GB:M62466; GB:M62467; GB:M62468; GB:M62469; GB:M62470; NID:g511867; PIDN:AAAA
R;Bornstein, P.; Alfli, D.; Devarayalu, S.; Framson, P.; Li, P.
J. Biol. Chem. 265, 16691-16698, 1990
A;Title: Characterization of the mouse thrombospondin gene and evaluation of the role
A;Reference number: A37905; MUID:90375546
A;Accession: A37905
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-490 <BOR>
A;Cross-references: GB:J05605; GB:J05606; NID:g201991; PIDN:AAA40431.1; PID:g554390
R;Laherty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seidin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A;Title: Characterization of mouse thrombospondin 2 sequence and expression during ce
A;Reference number: A42587; MUID:92147683
A;Accession: B42587
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1152, P',1154-1170 <LAH>
A;Cross-references: GB:M87276
A;Note: sequence extracted from NCBI backbone (NCBIP:81501)
R;Chen, H.; Aeschlimann, D.; Nowien, J.; Mosher, D.F.
FEBS Lett. 387, 36-41, 1996
A;Title: Expression and initial characterization of recombinant mouse thrombospondin
A;Reference number: S68787; MUID:96234006
A;Accession: S68787
A;Molecule type: protein
A;Residues: 19-26, 'X', 28-37 <CHE>
C;Complex: homotrimer, disulfide linked
C;Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;
C;Keywords: calcium binding; glycoprotein; homotrimer
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-1170/Product: thrombospondin 1 #status predicted <MAT>
F;317-375/Domain: von Willebrand factor type C repeat homology <VMC>
F;378-429/Domain: thrombospondin type 1 repeat homology <THR1>
F;434-490/Domain: thrombospondin type 1 repeat homology <THR2>
F;491-547/Domain: thrombospondin type 1 repeat homology <THR3>
F;551-586/Domain: EGF homology <EGF>
F;248,360,708,1067/Binding site: carboxydrate (Asn) (covalent) #status predicted

Query Match 53.6%; Score 59; DB 2; Length 1170;
Best Local Similarity 53.3%; Pred. No. 2.6;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 CANPSFVEGECPCSPC 18
| : : : ||||| |
Db 358 CSNATVPDGECCPRC 372

RESULT 11
S63986
collagen alpha 5 chain - sea urchin (Strongylocentrotus purpuratus) (fragment)
C;Species: Strongylocentrotus purpuratus (purple urchin)
C;Date: 20-Jul-1996 #sequence_revise 08-Nov-1996 #text_change 25-Apr-1997
C;Accession: S63986; S64638
R;Exposito, J.Y.; Boute, N.; Deleage, G.; Garrone, R.
Eur. J. Biochem. 234, 59-65, 1995
A;Title: Characterization of two genes coding for a similar four-cysteine motif of th
A;Reference number: S63985; MUID:96096722
A;Accession: S63986
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1376 <EXP>
A;Cross-references: EMBL:X89804
R;Exposito, J.Y.
submitted to the EMBL Data Library, July 1995
A;Reference number: S64637
A;Accession: S64638
A;Molecule type: DNA
A;Residues: 1-638, 'G', 660-870, 'G', 872-901, 'H', 903-1185, 'T', 1187-1214, 'Y', 1216-1376 <E
C;Genetics:
A;Gene: COLP5alpha

A: Introns: 73/1; 136/2; 221/1; 369/1; 517/1; 659/1; 799/1; 948/1; 1093/1; 1236/1
 C: Superfamily: von Willebrand factor type C repeat homology
 C: Keywords: extracellular matrix
 F: 15-73/Domain: von Willebrand factor type C repeat homology <WVC>

Query Match 53.6%; Score 59; DB 2; Length 1376;
 Best Local Similarity 61.1%; Pred. No. 2.9;
 Matches 11; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 PANCNPSFVEGCCPSC 18
 |||||
 Db 53 PAFCAEPIKPEGCCFLC 70

RESULT 12
 B41182
 collagen alpha 1(II) chain precursor (long splice form) - mouse
 C: Species: Mus musculus (house mouse)
 C: Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 16-Jul-1999
 C: Accession: B41182
 R: Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
 J. Biol. Chem. 266, 16862-16869, 1991
 A: Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
 A: Reference number: A41182; MUID: 91358489
 A: Accession: B41182
 A: Status: preliminary; not compared with conceptual translation
 A: Molecule type: DNA
 A: Residues: 1-1487 <NEW>
 A: Cross-references: GB:M65161
 C: Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
 C: Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
 F: 33-91/Domain: von Willebrand factor type C repeat homology <WVC>
 F: 1259-1487/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 53.6%; Score 59; DB 2; Length 1487;
 Best Local Similarity 60.0%; Pred. No. 3.1;
 Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 CANSFVGECCPSC 18
 |||||
 Db 74 CLNPEIPFGCCPIC 88

RESULT 13
 T21113
 hypothetical protein F19C6.3 - Caenorhabditis elegans
 C: Species: Caenorhabditis elegans
 C: Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C: Accession: T21113
 R: Harris, B.
 submitted to the EMBL Data Library, January 1995
 A: Reference number: Z19376
 A: Accession: T21113
 A: Status: preliminary; translated from GB/EMBL/DDBJ
 A: Molecule type: DNA
 A: Residues: 1-429 <WIL>
 A: Cross-references: EMBL:Z48006; PIDN:CAA88048.1; GSPDB:GN00028; CESP:F19C6.3
 A: Experimental source: clone F19C6
 C: Genetics:
 A: Gene: CESP:F19C6.3
 A: Map position: X
 A: Introns: 71/3; 137/2; 192/3; 217/3; 241/2; 273/3; 309/2; 366/3; 390/3

Query Match 53.2%; Score 58.5; DB 2; Length 429;
 Best Local Similarity 52.6%; Pred. No. 1.4;
 Matches 10; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 1 PANCNPSFVEGCCPSC 18
 |||||
 Db 252 PAVTCTHPMIRKDDCPSC 270

RESULT 14
 CGHU2V
 collagen alpha 2(V) chain precursor - human
 C: Species: Homo sapiens (man)
 C: Date: 31-Jul-1989 #sequence_revision 28-Jul-1995 #text_change 31-Dec-2000
 C: Accession: A31427; A54555; S43643; A25874; I55239; I59025; A25374; A30017
 R: Woodbury, D.; Benson-Chanda, V.; Ramirez, F.
 J. Biol. Chem. 264, 2735-2738, 1989
 A: Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the st
 A: Reference number: A31427; MUID: 89123368
 A: Accession: A31427
 A: Molecule type: mRNA
 A: Residues: 1-463 <WOO>
 A: Cross-references: GB:J04478; NID: g179697; PIDN: AAA51859.1; PID: g179698
 R: Greenspan, D.S.; Lee, S.T.; Lee, B.S.; Hoffman, G.G.
 Gene Expr. 1, 29-39, 1991
 A: Title: Homology between alpha2(V) and alpha1(III) collagen promoters and evidence f
 A: Reference number: A54555; MUID: 92314691
 A: Accession: A54555
 A: Molecule type: DNA
 A: Residues: 1-32 <GRE>
 A: Cross-references: GB:M58529; NID: g180834; PIDN: AAC41699.1; PID: g553235
 R: Moradi-Ameli, M.; Rousseau, J.C.; Kleman, J.P.; Champliand, M.F.; Boutillon, M.M.;
 Eur. J. Biochem. 221, 987-995, 1994
 A: Title: Diversity in the processing events at the N-terminus of type-V collagen.
 A: Reference number: S43642; MUID: 94237164
 A: Accession: S43643
 A: Molecule type: protein
 A: Residues: 288-291, 'P', 293-294, 'X', 296-297, 606, 'X', 608-617 <WOR>
 R: Weil, D.; Bernard, M.; Gargano, S.; Ramirez, F.
 Nucleic Acids Res. 15, 181-198, 1987
 A: Title: The pro alpha 2(V) collagen gene is evolutionarily related to the major fibr
 A: Reference number: A25874; MUID: 87146331
 A: Accession: A25874
 A: Molecule type: mRNA; DNA
 A: Residues: 398-1496 <WEL>
 A: Cross-references: GB:X04758; NID: g29588; PIDN: CAA28454.1; PID: g1340175
 R: Myers, J.C.; Loidl, H.R.; Stollé, C.A.; Seyer, J.M.
 J. Biol. Chem. 260, 5533-5541, 1985
 A: Title: Partial covalent structure of the human alpha 2 type V collagen chain.
 A: Reference number: I55239; MUID: 85182703
 A: Accession: I55239
 A: Status: translated from GB/EMBL/DDBJ
 A: Molecule type: mRNA
 A: Residues: 1002-1226 <RE2>
 A: Cross-references: GB:M10956; NID: g180427; PIDN: AAA52007.1; PID: g180428
 A: Note: part of this sequence were determined by protein sequencing
 R: Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
 A: Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long
 A: Reference number: I59025; MUID: 85216505
 A: Accession: I59025
 A: Status: translated from GB/EMBL/DDBJ
 A: Molecule type: mRNA
 A: Residues: 1003-1034 <RES>
 A: Cross-references: GB:M11135; NID: g179693; PIDN: AAA51857.1; PID: g179694
 A: Note: part of this sequence were determined by protein sequencing
 R: Myers, J.C.; Loidl, H.R.; Seyer, J.M.; Dion, A.S.
 J. Biol. Chem. 260, 11216-11222, 1985
 A: Title: Complete primary structure of the human alpha-2 type V procollagen COOH-term
 A: Reference number: A25374; MUID: 85289337
 A: Accession: A25374
 A: Molecule type: mRNA
 A: Residues: 1227-1417, 'T', 1419-1437, 'S', 1439-1496 <WYE>
 A: Cross-references: GB:M11718; NID: g180912; PIDN: AAA52058.1; PID: g180913
 R: Tsipouras, P.; Schwartz, R.C.; Liddell, A.C.; Salkeld, C.S.; Weil, D.; Ramirez, F.
 Genomics 3, 275-277, 1988
 A: Title: Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2, located

A:Reference number: A30017; MUID:89138450
A:Accession: A30017
A:Molecule type: DNA
A:Residues: 1449-1463; 'E', 1465-1495, 'A' <PSI>
A:Cross-references: GB:J03051; NID:g179695; PIDN:AAA51858.1; PID:g179696
A:Note: the authors translated the codon GAA for residue 1460 as Gln, and GAG for residue 1461 as Glu.
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit are 5-hydroxylated and subsequently O-glycosylated.
C:Comment: The amino-terminal propeptide domain appears not to be completely cleaved.
C:Genetics:
A:Gene: GDB:COL5A2
A:Cross-references: GDB:119064; OMIM:120190
A:Map position: 2q31-2q31
A:Introns: 33/1; 812/3; 830/3; 848/3; 884/3; 902/3; 922/3; 974/3; 1046/3; 1064/3; 1448/3
C:Complex: type V collagen may be a homotrimer of alpha 1(V) chains (see PIR:CGHUIV), a alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the length, is formed with desmosine cross-links made from lysine and allysine residues
C:Function:
A:Description: structural component of extracellular fibrous polymer associated with cell A:Note: may play a role in controlling the lateral growth of collagen I fibrils
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1250/Product: collagen alpha 2(V) chain #status predicted <MAT>
F:27-193/Domain: amino-terminal propeptide (uncleaved) #status predicted <NPP>
F:27-108/Region: nonhelical
F:40-99/Domain: von Willebrand factor type C repeat homology <VWC>
F:109-186/Region: helical
F:187-208/Region: nonhelical
F:209-1225/Region: helical
F:503-505/Region: cell attachment (R-G-D) motif
F:941-943/Region: cell attachment (R-G-D) motif
F:1064-1066/Region: cell attachment (R-G-D) motif
F:1067-1069/Region: cell attachment (R-G-D) motif
F:1097-1099/Region: cell attachment (R-G-D) motif
F:1124-1126/Region: cell attachment (R-G-D) motif
F:1133-1135/Region: cell attachment (R-G-D) motif
F:1225-1250/Region: carboxyl-terminal nonhelical telopeptide
F:1251-1496/Domain: carboxyl-terminal propeptide #status predicted <CPP>
F:1269-1496/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F:27/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F:193-194/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted
F:194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F:201/Modified site: allysine (Lys) #status predicted
F:290,293,296,608,614,1004,1007,1013,1028,1034/Modified site: 4-hydroxyproline (Pro) #status predicted
F:299,1139/Modified site: 5-hydroxylysine (Lys) #status predicted
F:299,1139/Binding site: carboxylate (Lys) (covalent) #status predicted
F:1025/Modified site: 5-hydroxylysine (Lys) #status experimental
F:1250-1251/cleavage site: Glu-Asp (procollagen C-endopeptidase) #status predicted
F:1259,1397/Binding site: carboxylate (Asn) (covalent) #status predicted
F:1293,1299,1325/Disulfide bonds: interchain #status predicted
F:1333-1494,1402-1447/Disulfide bonds: #status predicted

Query Match 51.8%; Score 57; DB 1; Length 1496;
Best Local Similarity 60.0%; Pred. No. 5.7;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 CANPSFVEGECPC 18
||| | | | | | | | |
DB 82 CADPVTTPPGCCPVC 96

RESULT 15
149562
alpha-1 type III collagen prepeptide - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C:Accession: I49562
R:Liau, G.; Yamada, Y.; de Crombrughe, B.
J. Biol. Chem. 260, 531-536, 1985
A:Title: Coordinate regulation of the levels of type III and type I collagen mRNA in mouse
A:Reference number: I49562; MUID:85080131

A:Accession: I49562
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-98 <RES>
A:Cross-references: GB:K03038; NID:g192273; PIDN:AAA37337.1; PID:g473632
C:Genetics:
A:Introns: 29/1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
F:32-92/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 50.9%; Score 56; DB 2; Length 98;
Best Local Similarity 50.0%; Pred. No. 0.9;
Matches 9; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 PATCANPSFVEGECPC 18
| | | | | | | | | |
DB 72 PLDCPNPEIPFGCCCAIC 89

Search completed: January 31, 2002, 09:24:52
Job time: 257 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:23:11 ; Search time 90.11 Seconds
(without alignments)
25.972 Million cell updates/sec

Title: US-09-536-087-6

Perfect score: 84

Sequence: 1 RESHFRGLQNVLVF 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_17.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	58.3	2828	Q9NR99	Q9nr99 homo sapien
2	44	52.4	229	Q28194	Q28194 bos taurus
3	44	52.4	376	Q15667	Q15667 homo sapien
4	44	52.4	616	Q9H906	Q9h906 homo sapien
5	43	51.2	350	Q9ZC67	Q9zc67 yersinia pe
6	43	51.2	1383	Q9SMR8	Q9smr8 arabidopsis
7	42	50.0	105	Q9B273	Q9b273 rhoptropus
8	42	50.0	385	Q9UKV6	Q9ukv6 homo sapien
9	42	50.0	468	Q9PIQ9	Q9piq9 homo sapien
10	42	50.0	838	Q9HDT4	Q9hdt4 trichoderma
11	41	48.8	117	Q9V5W1	Q9v5w1 drosophila
12	41	48.8	177	Q9HR28	Q9hr28 halobacteri
13	41	48.8	233	Q9CV47	Q9cv47 mus musculu
14	41	48.8	354	Q9CUG5	Q9cug5 mus musculu
15	41	48.8	383	Q9GUM2	Q9gum2 caenorhabdi
16	41	48.8	452	Q9GYL0	Q9gy10 caenorhabdi
17	41	48.8	464	Q9GY03	Q9gy03 mus musculu
18	41	48.8	516	Q9TWC9	Q9tmc9 amana eduli
19	41	48.8	693	Q23122	Q23122 arabidopsis

20	40	47.6	614	5	Q94287	Q94287 caenorhabdi
21	40	47.6	752	10	Q9LYU9	Q9lyu9 arabidopsis
22	40	47.6	4215	5	Q9W332	Q9w332 drosophila
23	39	46.4	201	2	Q9WZ24	Q9wz24 thermotoga
24	39	46.4	235	10	Q9FNJ3	Q9fnj3 arabidopsis
25	39	46.4	404	6	Q9MZE5	Q9mze5 sus scrofa
26	39	46.4	510	3	Q93999	Q93999 candida alb
27	39	46.4	568	3	P78947	P78947 schizosacch
28	39	46.4	605	3	Q9HGP1	Q9hgp1 schizosacch
29	39	46.4	865	3	Q43022	Q43022 schizosacch
30	39	46.4	1245	2	Q915W1	Q915w1 pseudomonas
31	39	46.4	1286	10	O23351	O23351 arabidopsis
32	38.5	45.8	136	5	O76687	O76687 caenorhabdi
33	38	45.2	99	12	Q9QKN9	Q9qkn9 hepatitis b
34	38	45.2	99	12	Q9QKN7	Q9qkn7 hepatitis b
35	38	45.2	99	12	Q9QKN3	Q9qkn3 hepatitis b
36	38	45.2	99	12	Q9QKN1	Q9qkn1 hepatitis b
37	38	45.2	99	12	Q9QKM9	Q9qkm9 hepatitis b
38	38	45.2	99	12	Q9QKM7	Q9qkm7 hepatitis b
39	38	45.2	99	12	Q9QKM5	Q9qkm5 hepatitis b
40	38	45.2	99	12	Q9QKM3	Q9qkm3 hepatitis b
41	38	45.2	99	12	Q9QKM1	Q9qkm1 hepatitis b
42	38	45.2	99	12	Q9QKL9	Q9qkl9 hepatitis b
43	38	45.2	99	12	Q9QKL7	Q9qkl7 hepatitis b
44	38	45.2	99	12	Q9QKK3	Q9qkk3 hepatitis b
45	38	45.2	185	2	P94932	P94932 mycoplasma

ALIGNMENTS

RESULT	1
Q9NR99	
ID	Q9NR99
AC	Q9NR99
DT	01-OCT-2000 (TREMBlrel. 15, Created)
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE	ADLICAN.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=PLACENTA;
RA	Crowl R.M., Luk D.;
RT	"Identification of the gene encoding Adlican, a novel protein expressed in human arthritic tissues."
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC	-1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX DOMAIN.
CC	EMBL; AF245505; AAF86402.1; -
DR	InterPro; IPR003598; Ig_c2.
DR	InterPro; IPR003600; Ig_like.
DR	InterPro; IPR003006; Ig_MHC.
DR	InterPro; IPR000483; LRR_Cterm.
DR	InterPro; IPR000372; LRR_Nterm.
DR	InterPro; IPR003592; LRR_out.
DR	InterPro; IPR003591; LRR_typ.
DR	Pfam; PF000047; Ig_12.
DR	Pfam; PF01463; LRRCT; 1.
DR	SMART; SM00408; IGC2; 10.
DR	SMART; SM00410; IG_like; 2.
DR	SMART; SM00370; LRR; 2.
DR	SMART; SM00082; LRRCT; 1.
DR	SMART; SM00013; LRRNT; 1.
DR	SMART; SM00369; LRR_Typ; 3.
SQ	SEQUENCE 2828 AA; 312292 MW; A18377D8554F1FE1 CRC64;

Query Match 58.3% Score 49; DB 4; Length 2828;
Best Local Similarity 57.1%; Pred. No. 8.6;

```
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 2 ESHFRGLQNVLHV 15
   :||:|||||
Db 1031 QSHLOGLTDNIHLV 1044

RESULT 2
Q28194
ID Q28194 PRELIMINARY; PRT; 229 AA.
AC Q28194;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE THROMBOSPONDIN-1 (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96331130; PubMed=8698834;
RA Lafeuillade B., Pellerin S., Keramidas M., Danik M., Chambaz E.M.,
RA Feige J.J.;
RT "Opposite regulation of thrombospondin-1 and corticotropin-induced
RT secreted protein/thrombospondin-2 expression by adrenocorticotrophic
RT hormone in adrenocortical cells."
RL J. Cell. Physiol. 167:164-172(1996).
DR EMBL; X89511; CAA61882.1; -.
DR InterPro; IPR001791; Laminin_G.
DR DR InterPro; IPR003129; TSPN.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SM00210; TSPN; 1.
FT NON_TER 1
FT 229
SQ SEQUENCE 229 AA; 25015 MW; 90D9EBC4E6B669C CRC64;

Query Match 52.4%; Score 44; DB 6; Length 229;
Best Local Similarity 61.5%; Pred. No. 4.6;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 4 HFRGLQNVLHV 16
   :||:|||||
Db 189 NFQGVQLNVRVF 201

RESULT 3
Q15667
ID Q15667 PRELIMINARY; PRT; 376 AA.
AC Q15667;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE THROMBOSPONDIN (TSP) PRECURSOR (FRAGMENT).
GN THBS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86287276; PubMed=3461443;
RA Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
RT "Characterization of a cDNA encoding the heparin and collagen binding
RT domains of human thrombospondin."
RL Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92278426; PubMed=1350660;
RA Burglin T.R., Barnes T.M.;
RT "Introns in sequence tags.";
```

```
RL Nature 357:367-368(1992).
DR EMBL; M14326; AAA61237.1; -.
DR InterPro; IPR001007; VWFC.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF00093; vwc; 1.
DR Pfam; PF02210; TSPN; 1.
DR PROSITE; PS01208; VWFC; 1.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
FT SIGNAL 1 18
FT CHAIN 19 >376
FT NON_TER 376
SQ SEQUENCE 376 AA; 41329 MW; 313B62C0D253BA2B CRC64;

Query Match 52.4%; Score 44; DB 4; Length 376;
Best Local Similarity 61.5%; Pred. No. 7.9;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 4 HFRGLQNVLHV 16
   :||:|||||
Db 207 NFQGVQLNVRVF 219

RESULT 4
Q9H9U6
ID Q9H9U6 PRELIMINARY; PRT; 616 AA.
AC Q9H9U6;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CDNA FLJ12543 FIS, CLONE NT2RM4000590, WEAKLY SIMILAR TO RING CANAL
DE PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
RA Wagatsuma M., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK022605; BAB14124.1; -.
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00651; BTB; 1.
DR Pfam; PF01344; Kelch; 5.
DR SMART; SM00225; BTB; 1.
DR PROSITE; PS50097; BTB; 1.
SQ SEQUENCE 616 AA; 69866 MW; CF904884FB34C80F CRC64;

Query Match 52.4%; Score 44; DB 4; Length 616;
Best Local Similarity 53.8%; Pred. No. 13;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 1 RESHFRGLQNVLHV 13
   :||:|||||
Db 224 REAHARQVLENIH 236

RESULT 5
Q9ZC67
ID Q9ZC67 PRELIMINARY; PRT; 350 AA.
AC Q9ZC67;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
```

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DE ARUF PROTEIN.
 GN ARUF.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=6/69;
 RA Buchrieser C., Rusnok C., Couve E., Frangeul L., Billault A.,
 RA Kunst F., Carniel E., Glaser P.;
 RT "DNA sequence of the 102 kbases unstable region of Yersinia pestis."
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL031866; CAA21340.1;
 SQ SEQUENCE 350 AA; 39467 MW; 1F3C02A1B04C92BD CRC64;

Query Match 51.2%; Score 43; DB 2; Length 350;
 Best Local Similarity 63.6%; Pred. No. 11;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 HFRGLQNVHL 14
 DB 305 HFRAILMNTL 315

RESULT 6
 Q9SMR8 PRELIMINARY; PRT; 1383 AA.
 AC Q9SMR8;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL 154.8 KDA PROTEIN.
 GN T5J17.20 OR AT4G39850.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
 RA Mayer K.F.X., Lemcke K., Schueller C.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL035708; CAB38898.1;
 DR EMBL: AL161596; CAB80648.1;
 DR InterPro: IPR003439; ABC transporter.
 DR InterPro: IPR001687; ATP GTP_A.
 DR Pfam: PF00005; ABC_tran_1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_2.
 KW Hypothetical protein.
 SQ SEQUENCE 1383 AA; 154829 MW; 56D03FB0521ECA09 CRC64;

Query Match 51.2%; Score 43; DB 10; Length 1383;
 Best Local Similarity 46.7%; Pred. No. 49;
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 ESHFRLQNVHLV 16
 DB 1336 DSHLKTILENRLV 1350

RESULT 7
 Q9B273 PRELIMINARY; PRT; 105 AA.
 AC Q9B273;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CYTOCHROME B (FRAGMENT).
 GN Cyte.
 OS Rhohtropus barnardi.
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Scieroglossa; Gekkota; Gekkonidae; Rhohtropus.
 OX NCBI_TaxID=152586;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lamb T., Bauer A.M.;
 RT "Mitochondrial phylogeny of Namib day geckos (Rhohtropus) based on
 RT cytochrome b and 16S rRNA sequences."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY026921; AAK13032.1;
 KW Mitochondrion.
 FT NON_TER 1
 FT NON_TER 105
 SQ SEQUENCE 105 AA; 11790 MW; FC7F680557BCB106 CRC64;

Query Match 50.0%; Score 42; DB 8; Length 105;
 Best Local Similarity 61.5%; Pred. No. 4.6;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RESHFRGLQNVH 13
 DB 67 REVHFGWLLRLH 79

RESULT 8
 Q9UKV6 PRELIMINARY; PRT; 385 AA.
 AC Q9UKV6;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE EROIL (FRAGMENT).
 GN EROIL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cabibbo A., Fabbri M., Pagani M., Sitia R.;
 RT "EROLL, a human homolog of the S. cerevisiae ERO1 gene."
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF123887; AAF06104.1;
 DR InterPro: IPR002048; EF-hand.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
 FT NON_TER 1
 SQ SEQUENCE 385 AA; 44719 MW; 24FCE8B4B5145D39 CRC64;

Query Match 50.0%; Score 42; DB 4; Length 385;
 Best Local Similarity 70.0%; Pred. No. 19;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 HFRGLQNVH 13
 DB 376 NFRNLLQNIH 385

```
RESULT 9
Q9P1Q9 PRELIMINARY; PRT; 468 AA.
AC Q9P1Q9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ER01-LIKE PROTEIN.
GN ER01-L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20138221; PubMed=10671517;
RA Cabibbo A., Pagani M., Fabbri M., Rocchi M., Farmery M.R.,
RA Bulleid N.J., Sitia R.;
RT "ER01-L, a human protein that favours disulfide bond formation in the
RT endoplasmic reticulum.";
RL J. Biol. Chem. 275:4827-4833(2000).
DR EMBL; AF081886; AAF35260.1; -.
DR InterPro; IPR002048; EF-hand.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
SQ SEQUENCE 468 AA; 54392 MW; 92CE6531C9CCA33 CRC64;

Query Match 50.0%; Score 42; DB 4; Length 468;
Best Local Similarity 70.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 HFRGLQNVH 13
Db 459 NFRNLQNIH 468

RESULT 10
Q9HDT4 PRELIMINARY; PRT; 838 AA.
AC Q9HDT4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE VESICLE FUSION FACTOR NSF1.
GN NSF1.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RU7C-30;
RA Saloheimo M.L.A., Wang H., Valkonen M., Vasara T., Huuskonen A.,
RA Riikonen M., Pakula T., Ward M., Penttilä M.;
RT "The secretory genes ypt1/yptA and nsf1/nsfA from the filamentous
RT fungi Trichoderma reesei and Aspergillus niger var. awamori: evidence
RT for global transcriptional regulation of the secretory pathway in T.
RT reesei.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ277109; CAC17745.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR003599; AAA_subfam.
DR Pfam; PF00004; AAA; 2.
DR SMART; SM00382; AAA; 2.
SQ SEQUENCE 838 AA; 90426 MW; 7C38E618771AA955 CRC64;

Query Match 50.0%; Score 42; DB 3; Length 838;
Best Local Similarity 72.7%; Pred. No. 43;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RESHFRGLQNL 11
Db 261 RESHARGILTN 271

RESULT 11
Q9V5W1 PRELIMINARY; PRT; 117 AA.
AC Q9V5W1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG9067 PROTEIN.
GN CG9067.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Fertaz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.W., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003826; AAF58685.1; -.
DR FlyBase; FBgn0033605; CG9067.
SQ SEQUENCE 117 AA; 13269 MW; 6D2CA9D0FAE00B42 CRC64;

Query Match 48.8%; Score 41; DB 5; Length 117;
Best Local Similarity 37.5%; Pred. No. 7.8;
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 RESHFRGLQNVH 16
Db 70 RENEVRAIFRNHL 85
```

0;

QY 1 RESHFRGLLQ 11
 ||:|||||
 Db 224 REAHFRKLQ 234

RESULT 15

Q9GUM2 PRELIMINARY; PRT; 383 AA.
 AC Q9GUM2;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE HYPOTHETICAL PROTEIN Y73E7A.B.
 GN Y73E7A.B.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AC025727; AAC23384.1; -;
 DR InterPro; IPR001998; Xylose.isom.
 DR PROSITE; PS00173; XYLOSE_ISOMERASE_2; UNKNOWN1.
 SQ SEQUENCE 383 AA; 43913 MW; ADEDF44275B2CE81 CRC64;

Query Match 48.8%; Score 41; DB 5; Length 383;
 Best Local Similarity 53.8%; Pred. No. 28;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 RESHFRGLLQNVH 13
 ||:|||||
 Db 159 REAHLRLMLNLH 171

Search completed: January 31, 2002, 09:36:13
 Job time: 782 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:23:04 ; Search time 96.28 Seconds
(without alignments)
13.848 Million cell updates

Title: US-09-536-087-7

Perfect score:

Sequence: 1 PATCANPSFVEGECCPSC 18

Scoring table: BLOSUM62

scoring table. BBS0M02
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

rose processing. Minimum Match 0% Maximum Match 100%

Maximum Match 100%
Listing first 45 summaries

Database :

```
Database : A_Geneseq_1101:*
1: /SID58/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SID58/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SID58/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SID58/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SID58/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SID58/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SID58/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SID58/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SID58/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SID58/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SID58/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SID58/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SID58/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SID58/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SID58/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SID58/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	110	100.0		18	21	AAB19680	Human thrombospond
2	105	95.5		1172	21	AAB19677	Human thrombospond
3	105	95.5		1172	21	AAB00043	Human thrombospond
4	68	61.8		322	21	AAB40954	Human OREF ORF718
5	68	61.8		732	22	AAB61140	Human NOV10 protein
6	68	61.8		1036	21	AAE82776	Human chordin rela
7	68	61.8		1036	21	AAV53034	Human secreted pro
8	68	61.8		1036	22	AAU07141	Human CRIM1 protein
9	68	61.8		1036	22	AAU12242	Human PR04330 poly
10	68	61.8		1037	22	AAU07142	Mouse CRIM1 protein
11	68	61.8		1048	22	AAU07143	Chicken CRIM1 prot

Type I, p80 IL-1-r
Human Interleukin-1
Human colorectal cancer
Human gene 19 encodes
Melanoma associated
Human p53 target mRNA
C. elegans CRIMI protein
Human chordin related
Human secreted protein
Human net-related
Peptide #2360 encodes
Peptide #2446 encodes
Peptide #2344 encodes
Human colon cancer
Human TSP1 protein
Thrombospondin I
Thrombospondin I
Human thrombospondin
Human concatamers
Thrombospondin I
Angiotensin convert
Human cancer associated
Angiotensin convert
Angiotensin convert
Angiotensin convert
Angiotensin convert
Human thrombospondin
Human shear stress
Human variant thrombospondin
Rat type II collagen
Human secreted protein
Human EXMAD-10 sequence
Human adipose tissue
Gene 16 human secreted
Human type II collagen
Human type I collagen
Xenopus frog proteoglycan

ALIGNMENTS

RESULT	1
AAB19680	
ID	AAB19680 standard; Peptide; 18 AA.
XX	
XX	
AC	AAB19680;
AC	
XX	
XX	
DT	05-FEB-2001 (first entry)
XX	
XX	
DE	Human thrombospondin-2 peptide 2 (aa357-374)

RESULT 1

AAB19680
ID AAB19680 standard; Peptide; 18 AA.

AAC 19680:

DT 05-FEB-2001 (first entry)

Human thrombospondin-2 peptide 2 (aa357-374).

Thrombospondin-2; TSP-2; human; human; angiogenesis; cell proliferation; KW

KW melanoma; tumour; cancer; squamous cell carcinoma; antiangiogenic;

KW prostate cancer; psoriasis; rosecea dermatosis; antitumour;

KW therapy.

OS Homo sapiens.

PN WO200057899-A1.

PD 05-OCT-2000.

PF 24-MAR-2000; 2000WO-US07835.

PR 31-MAR-1999; 99US-0127221.

PA (GEHO) GEN HOSPITAL CORP.

PI Detmar M, Streit M;

WPI: 2000-656131/63.

PT Treating a disorder characterized by unwanted cell proliferation e.g.
PT precancerous, cancerous or neoplastic cells or presence of tumor
PT preferably of skin or prostate, comprises increasing thrombospondin-2

PT activity -
 PS Disclosure; Page 40; 73pp; English.
 CC The present sequence is that of peptide 2, derived from the
 CC procollagen domain (amino acids 357-374) of human thrombospondin-2
 CC (TSP-2, see AAB19677). The peptide is 1 of 5 synthetic peptides (see
 CC AAB19679-83) used with human dermal microvascular endothelial cells
 CC (HMEC) to determine the effect of TSP-2 on endothelial cell
 CC migration. Peptide 2 did not significantly modify HMEC migration,
 CC in contrast to peptide 7 (see AAB19683) from a type 1 repeat of
 CC TSP-2, which inhibited migration. The invention is based on the
 CC discovery that overexpression of TSP-2 decreases tumor size in
 CC vivo, and features methods for modulating unwanted angiogenesis and
 CC cell proliferation by increasing TSP-2 activity.
 XX Sequence 18 AA;
 SQ
 Query Match 100.0%; Score 110; DB 21; Length 18;
 Best Local Similarity 100.0%; Pred. No. 3.5e-07;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 PATCANPSFVEGECPC 18
 Db 1 patcanspfvegeccpc 18
 |||||
 RESULT 2
 AAB19677
 ID AAB19677 standard; Protein: 1172 AA.
 AC AAB19677;
 DT 05-FEB-2001 (first entry)
 DE Human thrombospondin-2.
 KW Thrombospondin-2; TSP-2; human; angiogenesis; cell proliferation;
 KW melanoma; tumor; cancer; squamous cell carcinoma; antiangiogenic;
 KW prostate cancer; psoriasis; rosecea dermatosis; antitumor;
 KW therapy.
 XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX Region 382..429
 FT /note= "type 1 repeat"
 FT Region 384..390
 FT /note= "WSPWAEW sequence involved in antiangiogenic
 FT activity"
 FT Region 438..490
 FT /note= "type 1 repeat"
 FT Region 495..547
 FT /note= "type 1 repeat"
 XX
 XX WO200057899-A1.
 XX
 XX 05-OCT-2000.
 XX
 XX 24-MAR-2000; 2000WO-US07835.
 XX
 XX 31-MAR-1999; 99US-0127221.
 XX
 XX (GEHO) GEN HOSPITAL CORP.
 XX
 XX Detmar M, Streit M;
 XX
 XX WPI: 2000-656131/63.
 XX N-PSDB: AAA88669.
 XX
 XX Treating a disorder characterized by unwanted cell proliferation e.g.
 XX precancerous, cancerous or neoplastic cells or presence of tumor

PT preferably of skin or prostate, comprises increasing thrombospondin-2
 PT activity -
 XX Disclosure; Fig 2; 73pp; English.
 PS
 CC The present sequence is that of human thrombospondin-2 (TSP-2).
 CC The invention is based on the discovery that overexpression of
 CC TSP-2 decreases tumor size in vivo, and features methods for
 CC modulating unwanted angiogenesis and tumor growth. Treatment of
 CC unwanted cell proliferation or angiogenesis involves increasing
 CC TSP-2 activity. This is achieved by administering an agent which
 CC increases TSP-2 activity, especially a TSP-2 polypeptide, a TSP-2
 CC derived polypeptide or retro-inverso peptide, a nucleic acid
 CC encoding TSP-2, an agonist of TSP-2, or an agent that increases
 CC TSP-2 gene expression. The TSP-2 polypeptide may include at least
 CC one type 1 repeat such as the WSPWAEW peptide (see AAB19683),
 CC which is involved in the antiangiogenic activity of TSP-2. The
 CC method is used to treat a disorder characterised by pre-cancerous,
 CC cancerous or neoplastic cells, or the presence of a tumour, or a
 CC skin cell proliferation. Such disorders include malignant
 CC melanoma, prostate cancer, squamous cell carcinoma, aged skin,
 CC rosecea dermatosis, psoriasis, and skin damage caused by
 CC photoradiation (all claimed). Evaluating the presence of TSP-2
 CC nucleic acid or protein is useful for diagnosing a subject at risk
 CC of unwanted cell proliferation or angiogenesis. Methods are also
 CC provided of identifying compounds that modulate TSP-2 activity.
 XX Sequence 1172 AA;
 SQ
 Query Match 95.5%; Score 105; DB 21; Length 1172;
 Best Local Similarity 94.4%; Pred. No. 5.8e-05;
 Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 PATCANPSFVEGECPC 18
 Db 357 patcanspfvegeccpc 374
 |||||
 RESULT 3
 AAB00043
 ID AAB00043 standard; Protein: 1172 AA.
 AC AAB00043;
 XX
 XX 08-NOV-2000 (first entry)
 XX
 XX Human thrombospondin-2 (TSP-2).
 KW TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
 KW thrombospondin; angiogenesis; tumor; treatment; cancer;
 KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
 KW glaucoma.
 XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX Region 381..436
 FT /label= Type 1 repeat region
 FT Region 437..493
 FT /label= Type 1 repeat region
 FT Region 494..550
 FT /label= Type 1 repeat region
 XX
 XX WO20004908-A2.
 XX
 XX 03-AUG-2000.
 XX
 XX 01-FEB-2000; 2000WO-US02482.
 XX
 XX 01-FEB-1999; 99US-0118053.
 XX

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 XX Lawler JW;
 PI WPI; 2000-514823/46.
 XX
 XX Nucleic acids encoding chimeric proteins such as cartilage oligomeric
 PT matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for
 PT inhibiting angiogenesis and treating diseases such as cancer
 XX
 XX Disclosure; Fig 2; 40pp; English.
 PS
 XX New nucleic acids are described which encode a protein comprising
 CC the second and third type 1 repeats of human TSP (thrombospondin)-1,
 CC but not the TGF (transforming growth factor)-beta activation region
 CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing
 CC the second and third type-1 repeats and the COMP (cartilage
 CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was
 CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1
 CC caused inhibition of the growth of tumours in mice models.
 CC Thus the nucleic acids and proteins may be useful for treating
 CC angiogenesis related diseases such as cancer (by reducing the rate of
 CC growth and size of tumours), arthritis, psoriasis, diabetic
 CC retinopathy, corneal graft rejection, and glaucoma. They may also be
 CC used for treating human immunodeficiency virus (HIV) infection.
 CC Anti-angiogenic therapy has little toxicity, does not require the
 CC therapeutic agent to enter tumour cells or cross the blood-brain
 CC barrier, controls tumour growth independently of growth of
 CC tumour cell heterogeneity, and does not induce drug resistance.
 XX
 XX Sequence 1172 AA;
 SQ
 Query Match 95.5%; Score 105; DB 21; Length 1172;
 Best Local Similarity 94.4%; Pred. No. 5.8e-05;
 Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PATCANPSFVEGECPCSC 18
 Db 357 patcaspsfvegeccpsc 374
 |||||:|||||||
 |||||:|||||||
 RESULT 4
 AAB40954
 ID AAB40954 standard; Protein; 322 AA.
 AC AAB40954;
 XX
 DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF718 polypeptide sequence SEQ ID NO:1436.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiprosiatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX

PF 31-MAR-2000; 2000WO-US08621.
 XX
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 03-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 XX Shimkets RA, Leach M;
 XX
 XX WPI; 2000-602362/57.
 DR N-PSDB; AAC75163.
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 XX Claim 11; Page 1214-1215; 5507pp; English.
 PS
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiprosiatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 XX Sequence 322 AA;
 SQ
 Query Match 61.8%; Score 68; DB 21; Length 322;
 Best Local Similarity 68.8%; Pred. No. 0.54;
 Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 3 TCANPSFVEGECPCSC 18
 Db 147 tctnpvkpgccpcvc 162
 |||||:|||||
 |||||:|||||
 RESULT 5
 AAB61140
 ID AAB61140 standard; Protein; 732 AA.
 XX
 AC AAB61140;
 XX
 DT 30-MAR-2001 (first entry)
 DE Human NOV10 protein.
 XX
 KW Human; NOVX; antinflammatory; cytostatic; neuroprotective;
 KW cerebroprotective; immunomodulator; vulnary; vasotropic; gene therapy;
 KW hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;
 KW diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200075321-A2.
 XX

PD 14-DEC-2000.
 PF 01-JUN-2000; 2000WO-US15303.
 XX
 XX 03-JUN-1999; 99US-0137322.
 PR 16-MAR-2000; 2000US-0189810.
 PR 22-MAR-2000; 2000US-0191158.
 PR 30-MAR-2000; 2000US-0193086.
 PR 31-MAY-2000; 2000US-0137322.
 XX
 XX (CURA-) CURAGEN CORP.

PI Shinkets RA, Fernandes E, Herrman J, Vernet C;
 XX
 DR WPI; 2001-102403/11.
 DR N-PSDB; AAF27858.

XX New NOVX polypeptides and polynucleotides, useful in gene therapy, as a
 PT diagnostic marker, protein therapeutic, antibody or small molecule drug
 PT target for treating immune, proliferative and metabolic diseases and
 PT wound healing

PS Claim 1; Page 39-42; 194pp; English.

XX The present sequence is a new isolated polypeptide (NOVX). The NOVX
 CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
 CC treating or preventing NOVX-associated disorders. They are also useful
 CC for determining the presence of or a predisposition to a disease
 CC associated with altered levels of the NOVX polypeptide or nucleic acid.
 CC These NOVX-associated disorders include hyperplasias, tumours,
 CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
 CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
 CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
 CC are especially useful in gene therapy. Specifically, NOVX is useful as
 CC a diagnostic marker or prognostic marker, protein therapeutic and
 CC antibody target or small molecule drug target to treat disorders in the
 CC immune response pathway, thyroid and metabolic diseases, bone metabolic
 CC disorders, diseases of the pancreas (e.g. diabetes or digestive
 CC disorders), proliferative diseases, or tissue regeneration and
 CC development (e.g. wound healing or treatment of burns).

SQ Sequence 732 AA;

Query Match 61.8%; Score 68; DB 22; Length 732;
 Best Local Similarity 68.8%; Pred. No. 1.1;
 Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 TCANPSFVEGECPCSPC 18
 ||||| ||||| |
 DB 283 tctnpkvpgeccpvc 298

RESULT 6

RAY82776
 ID AAY82776 standard; Protein; 1036 AA.

XX AAY82776;

DT 19-JUN-2000 (first entry)

XX Human chordin related protein (Clone dj167_19).

XX Chordin related protein; cartilage; bone; connective tissue;
 KW periodontal disease; osteoporosis; burn; incision; ulcer; neuron;
 KW astrocyte; glial cell; transplantation; nerve; epidermis; muscle;
 KW liver; brain; lung; cardiac; pancreas; kidney; growth;
 KW differentiation; TGF-Beta; angiogenesis; chemotaxis;
 KW chemoattraction; collagen synthesis; fibrosis; cell adhesion;
 KW cell migration; fertility; reproduction; haematopoiesis;
 KW erythroid cell; tumour; dietary supplement; growth medium.

OS Homo sapiens.

XX WO200009551-A1.

XX 24-FEB-2000.

XX 10-AUG-1999; 99WO-US18117.

XX 10-AUG-1998; 98US-0095880.

XX 06-MAY-1999; 99US-0306111.

XX (GEMY) GENETICS INST INC.

XX Jacobs K, McCoy JM, Lavallie ER, Collins-racie LA, Merberg D;
 PI Treacy M, Diblasio-smith E, Widom A;

XX WPI; 2000-205978/18.

XX N-PSDB; AAZ93172.

XX New polynucleotides encoding secreted human proteins, useful for
 PT treating e.g. broken bones, craniofacial defects, periodontal disease,
 PT osteoporosis, burns, incisions or ulcers

PS Claim 21; Page 94-98; 105pp; English.

XX The human chordin related protein and polynucleotides encoding them
 CC are predicted to have biological activities which would make them
 CC suitable for treating, preventing or ameliorating medical conditions
 CC which involve defects in cartilage, bone or connective tissue
 CC formation and damage to cartilage, bone or connective tissue, e.g.
 CC broken bones, congenital, trauma-induced, or
 CC oncologic-resection-induced craniofacial defects, periodontal
 CC disease, defects in the periodontal ligament or attachment apparatus,
 CC damage to the periodontal ligament or attachment apparatus,
 CC osteoporosis, burns, incisions or ulcers. The proteins may also
 CC affect neuronal, astrocytic, and glial cell survival and therefore be
 CC useful in transplantation and treatment of conditions exhibiting a
 CC decrease in neuronal survival and repair. The proteins may also be
 CC useful for the treatment of conditions related to other types of
 CC tissue, such as nerve, epidermis, muscle, and other organs such as
 CC liver, brain, lung, cardiac, pancreas, and kidney tissue. The
 CC proteins may further be useful for the treatment of relatively
 CC undifferentiated cell populations, such as embryonic cells, or stem
 CC cells, to enhance growth and/or differentiation of the cells.
 CC The proteins may also have other useful properties characteristic of
 CC the TGF-beta superfamily of proteins. Such properties include
 CC angiogenic, chemotactic, and/or chemoattractant properties, and
 CC effects on cells including induction or inhibition of collagen
 CC synthesis, fibrosis, differentiation responses, cell proliferative
 CC responses, and responses involving cell adhesion, migration, and
 CC extracellular matrices. These properties make the proteins potential
 CC agents for wound healing, reduction of fibrosis, and reduction of
 CC scar tissue formation. Chordin-related proteins may also be useful
 CC for advancement of the onset of fertility in sexually immature
 CC mammals, so as to increase the lifetime reproductive performance of
 CC domestic animals such as cows, sheep and pigs. Chordin-related
 CC proteins may also be useful in modulating hematopoiesis by inducing
 CC the differentiation of erythroid cells, for suppressing the
 CC development of gonadal tumors, or for augmenting the activity of
 CC BMPs. The proteins may also have value as a dietary supplement, or
 CC as a component of cell culture media.

XX Sequence 1036 AA;

Query Match 61.8%; Score 68; DB 21; Length 1036;

Best Local Similarity 68.8%; Pred. No. 1.5;

Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 TCANPSFVEGECPCSPC 18

||||| ||||| |

DB 441 tctnpkvpgeccpvc 456

```
RESULT 7
AAY53034
ID AAY53034 standard; Protein; 1036 AA.
XX AC
XX AC AAY53034;
XX DT 29-FEB-2000 (first entry)
XX DE Human secreted protein clone dj167_19 protein sequence SEQ ID NO:74.
XX KW Human; secreted protein; nutritional; cytokine; cell proliferation;
XX KW differentiation; immune stimulating; vaccine; suppression;
XX KW haematopoiesis regulation; tissue growth; activin; inhibin;
XX KW chemotactic; chemokinetic; haemostatic; thrombolytic; receptor;
XX KW ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
XX KW tumour inhibition; gene therapy.
XX OS Homo sapiens.
XX PN WO9957132-A1.
XX PD 11-NOV-1999.
XX PF 07-MAY-1999; 99WO-US09970.
XX PR 07-MAY-1998; 98US-0084564.
XX PR 02-JUN-1998; 98US-0087645.
XX PR 22-JUL-1998; 98US-0093712.
XX PR 31-JUL-1998; 98US-0094935.
XX PR 10-AUG-1998; 98US-0095880.
XX PR 11-AUG-1998; 98US-0096068.
XX PR 06-MAY-1999; 99US-0096068.
XX PA (GEMY ) GENETICS INST INC.
XX PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Bowman MR;
XX PI DiBlasio-Smith E, Widom A;
XX WPI: 2000-052937/04.
XX DR N-PSDB; AA233352.
XX N New polynucleotides encoding secreted human proteins, derived from
XX adult placenta, adult retina, fetal brain, fetal
XX Claim 83; Page 426-429; 492pp; English.
XX CC The present invention describes new human secreted proteins which were
XX isolated from adult placenta, adult retina, foetal brain, foetal kidney,
XX adult blood, adult brain, adult thyroid, adult bladder, adult neural
XX tissue, adult testes, and adult lymph node cDNA libraries. The human
XX secreted proteins, and the polynucleotides encoding them, are predicted
XX to have biological activities which would make them suitable for
XX treating, preventing or ameliorating medical conditions in humans and
XX animals. Suggested activities include nutritional activity, cytokine
XX and cell proliferation/differentiation activity, immune stimulating
XX (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
XX activity, tissue growth activity, activin/inhibin activity,
XX chemotactic/chemokinetic activity, haemostatic and thrombolytic
XX activity, receptor/ligand activity, anti-inflammatory activity,
XX cadherin/tumour invasion suppressor activity, and tumour inhibition
XX activity. The polynucleotides are also stated to be useful for gene
XX therapy. AA233316 to AA233373 encode human secreted proteins, and
XX AA52998 to AA53060 represent human secreted proteins, given in the
XX present invention.
XX SQ Sequence 1036 AA;

Oy 3 TCAMPSFVEGECPPSC 18
Db 441 tctnpkvpgccpvc 456

RESULT 8
AAU07141
ID AAU07141 standard; Protein; 1036 AA.
XX AC AAU07141;
XX DT 24-OCT-2001 (first entry)
XX DE Human CRIM1 protein.
XX KW CRIM-1; Human; human chromosome 2p21-16.3; ophthalmological;
XX KW neuroprotective; renal; osteopathic; dental; vulnery; immunogen;
XX KW antibody; gene therapy; neurodegenerative disease; eye disorder;
XX KW cataract; bone morphogenic protein; BMG; renal disease; bone abnormality;
XX KW tooth abnormality; wound; S52.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Peptide 1..17
XX FT /label= Signal_peptide
XX FT Domain 1..901
XX FT /label= Ectodomain
XX FT /note= "This sequence is specifically claimed in claim
XX FT 15"
XX FT Protein 18..1036
XX FT /label= Mature_CRIM1
XX FT Region 200..207
XX FT /note= "Conserved N-terminal motif"
XX FT Region 336..391
XX FT /label= CR_1
XX FT /note= "Cysteine rich repeat"
XX FT Region 403..456
XX FT /label= CR_2
XX FT /note= "Cysteine rich repeat"
XX FT Misc-difference 414
XX FT /note= "Encoded by GAC"
XX FT Region 608..662
XX FT /label= CR_3
XX FT /note= "Cysteine rich repeat"
XX FT Region 679..734
XX FT /label= CR_4
XX FT /note= "Cysteine rich repeat"
XX FT Region 753..808
XX FT /label= CR_5
XX FT /note= "Cysteine rich repeat"
XX FT Region 819..873
XX FT /label= CR_6
XX FT /note= "Cysteine rich repeat"
XX PN WO200138519-A1.
XX DT 31-MAY-2001.
XX PR 24-NOV-2000; 2000WO-AU01435.
XX PR 26-NOV-1999; 99AU-0004348.
XX PA (UYQU ) UNIV QUEENSLAND.
XX PI Little M, Yamada T, Holmes G, Georgas K, Kolle G, Wilkinson L;
XX WPI: 2001-343951/36.
XX DR N-PSDB; AAS11601.
XX N Nucleic acids from human chromosome 2p21-16.3 and the encoded peptide,
XX useful for preventing, diagnosing and treating e.g. eye disease,
XX especially cataract formation -
```

XX Claim 11; Fig 1; 169pp; English.
 XX The invention relates to nucleic acids from human chromosome 2p21-16.3
 CC and the encoded peptide (and mouse and chicken orthologues) that
 CC comprises a RGECCPLP group, an insulin-like growth factor binding protein
 CC (IGFBP)-like domain, cysteine-rich domains, an RGD (undefined) group
 CC and a transmembrane domain. The protein, e.g. CRIM1, interacts with
 CC peptides of the transforming growth factor superfamily. A composition
 CC comprising an expression construct comprising the nucleic acids of the
 CC invention or a mimetic which antagonises or mimics an activity of a CRIM1
 CC polypeptide may be used in a method for modulating the biological
 CC activity of a polypeptide of the bone morphogenic protein (BMP) family.
 CC In this way they may be used to prevent or treat an eye disease,
 CC especially cataract formation. They may also be used to treat
 CC neurodegenerative diseases, renal and kidney disease, bone and tooth
 CC abnormalities, wounds and skin damage, e.g. by use of the nucleic acid in
 CC gene therapy by using antibodies directed against CRIM1 polypeptides.
 CC The present sequence represents human CRIM1 (AKA S52).
 XX
 SQ Sequence 1036 AA;

Query Match 61.8%; Score 68; DB 22; Length 1036;
 Best Local Similarity 68.8%; Pred. No. 1.5;
 Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 3 TCANPSFVEGECPC 18
 DB 441 tctnpkvpgeccpvc 456
 || || | ||||| |

RESULT 9
 AAU12242
 ID AAU12242 standard; Protein; 1036 AA.
 AC AAU12242;
 XX
 DT 24-OCT-2001 (first entry)
 DE Human PRO4330 polypeptide sequence.
 XX
 KW Human secretory and transmembrane; PRO; mammalian; cancer; lung;
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
 KW adipocyte; A-peptide; factor VIIA; gene therapy.
 XX
 OS Homo sapiens.

XX WO200140466-A2.
 XX
 PN 07-JUN-2001.
 PD
 XX
 PF 01-DEC-2000; 2000WO-US32678.
 XX
 PR 01-DEC-1999; 99WO-US28301.
 PR 01-DEC-1999; 99WO-US28634.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28564.
 PR 02-DEC-1999; 99WO-US28565.
 PR 09-DEC-1999; 99US-0170262.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 20-DEC-1999; 99WO-US30999.
 PR 30-DEC-1999; 99WO-US31243.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 06-JAN-2000; 2000WO-US00376.
 PR 11-FEB-2000; 2000WO-US03565.
 PR 18-FEB-2000; 2000WO-US04341.
 PR 18-FEB-2000; 2000WO-US04342.
 PR 22-FEB-2000; 2000WO-US04414.
 PR 24-FEB-2000; 2000WO-US04914.
 PR 24-FEB-2000; 2000WO-US05004.

PR 01-MAR-2000; 2000WO-US05601.
 PR 20-MAR-2000; 2000WO-US07377.
 PR 21-MAR-2000; 2000WO-US07532.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 10-NOV-2000; 2000WO-US30873.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
 XX
 DR WPI: 2001-408281/43.
 DR N-PSDB; AAS21314.
 XX
 PT Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical
 XX
 PS Claim 12; Fig 142; 813pp; English.

XX AAU12172-AAU12446 represent novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.
 XX
 SQ Sequence 1036 AA;

Query Match 61.8%; Score 68; DB 22; Length 1036;
 Best Local Similarity 68.8%; Pred. No. 1.5;
 Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 TCANPSFVEGECPC 18
 DB 441 tctnpkvpgeccpvc 456
 || || | ||||| |
 RESULT 10
 AAU07142
 ID AAU07142 standard; Protein; 1037 AA.

XX AAU07142;
 XX
 DT 24-OCT-2001 (first entry)
 DE Mouse CRIM1 protein.
 XX
 KW CRIM-1; Mouse; human chromosome 2p21-16.3; ophthalmological;
 KW neuroprotective; renal; osteopathic; dental; vulnery; immunogen;
 KW antibody; gene therapy; neurodegenerative disease; eye disorder;
 KW cataract; bone morphogenic protein; BMG; renal disease; bone abnormality;

CC peptides of the transforming growth factor superfamily. A composition
 CC comprising an expression construct comprising the nucleic acids of the
 CC invention or a mimetic which antagonises or mimics an activity of a CRIM1
 CC polypeptide may be used in a method for modulating the biological
 CC activity of a polypeptide of the bone morphogenic protein (BMP) family.
 CC In this way they may be used to prevent or treat an eye disease,
 CC especially cataract formation. They may also be used to treat
 CC neurodegenerative diseases, renal and kidney disease, bone and tooth
 CC abnormalities, wounds and skin damage, e.g. by use of the nucleic acid in
 CC gene therapy by using antibodies directed against CRIM1 polypeptides.
 CC The present sequence represents chicken CRIM1 (AKA S52).
 XX
 SQ Sequence 1048 AA;

Query Match 61.8%; Score 68; DB 22; Length 1048;
 Best Local Similarity 61.1%; Pred. No. 1.6;
 Matches 11; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 PATCANPSFVEGECPC 18
 I I I I I I I I I
 Db 868 plpcaeipnvegscpmc 885

RESULT 12

AAW19989
 ID AAW19989 standard; Protein; 176 AA.

AC AAW19989;

DT 27-AUG-1997 (first entry)

DE Type I, p80 IL-1-receptor intracellular domain ligand.

KW IL; interleukin; receptor; ligand; screening assay; inhibitor;
 KW IL-1 mediated response; inflammation; inflammatory; antibody;
 KW intracellular domain; CANP; calcium activated neutral protease.

XX Homo sapiens.

XX WO9640507-A1.

XX 19-DEC-1996.

XX 06-MAY-1996; 96WO-US06363.

XX 07-JUN-1995; 95US-0487942.

XX (GEMY) GENETICS INST INC.

XX Graham J, Lin L;

XX WPI; 1997-052315/05.

DR N-PSDB; AAT71216.

XX Interleukin-1 receptor intracellular ligand proteins and related DNA
 PT - used to identify inhibitors of the proteins for treatment of
 PT inflammation

XX Claim 2; Page 30; 54pp; English.

XX AAW19989 represents a protein that has interleukin-1 receptor (IL-1-R)
 CC intracellular ligand activity. IL-1-R intracellular ligand proteins
 CC are used to screen for agents (e.g. antibodies) that are capable of
 CC inhibiting or blocking the binding of an IL-1-R intracellular ligand
 CC to the intracellular domain of IL-1-R, i.e. inhibitors of IL-1
 CC activity. Such agents can be used to treat inflammatory conditions.

XX Sequence 176 AA;

Query Match 60.9%; Score 67; DB 18; Length 176;
 Best Local Similarity 61.1%; Pred. No. 0.42;

Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 Qy 1 PATCANPSFVEGECPC 18
 I I I I I : I I I I
 Db 150 patcavpnipgacpcvc 167

RESULT 13

AAAB37794
 ID AAB37794 standard; Protein; 176 AA.

XX AAB37794;

XX 23-FEB-2001 (first entry)

XX Human interleukin-1 receptor intracellular ligand protein #1.

XX Human; interleukin-1; IL-1; IL-lalpha; IL-lbeta; IL-1 receptor;
 KW antiinflammatory; haemostatic; antibacterial; immunosuppressive;
 KW immunomodulator; cardiant; cytostatic; neuroprotective; respiratory;
 KW inflammation; infection; sepsis; cachexia; autoimmune disorder;
 KW cardiovascular disorder; chronic myelogenous leukaemia;
 KW multiple sclerosis; inflammatory bowel disease; Crohn's disease.

XX Homo sapiens.

XX WO200064479-A1.

XX 02-NOV-2000.

XX 26-APR-2000; 2000WO-US11700.

XX 27-APR-1999; 99US-0301274.

XX (ANTI-) ANTIBODY SYSTEMS INC.

XX Fredeking TM, Ignatyev GM;

XX WPI; 2000-679646/66.

XX Novel compositions comprising tetracycline or tetracycline-like
 PT compounds for the treatment and/or prevention of acute inflammatory
 PT responses and diseases, e.g. septic shock and immune complex-induced
 PT colitis -

XX Disclosure; Page 155-156; 183pp; English.

XX The present sequence is given in a specification relating to novel
 CC compositions and methods containing tetracycline or tetracycline-like
 CC compounds for treating and/or preventing acute inflammatory responses and
 CC diseases. Such diseases include acute inflammatory conditions associated
 CC with viral haemorrhagic diseases (including diseases caused by
 CC Bunyaviridae, Filoviridae, Flaviviridae or Arenaviridae viruses),
 CC parasitic diseases, bacterial infections, sepsis, cachexia, autoimmune
 CC disorders, acute cardiovascular events, chronic myelogenous leukaemia and
 CC transplanted bone marrow-induced graft-versus-host disease, septic shock,
 CC immune complex-induced colitis, cerebrospinal fluid inflammation,
 CC multiple sclerosis, inflammatory responses associated with trauma,
 CC systemic inflammatory response syndrome (SIRS), adult respiratory
 CC distress syndrome (ARDS), acute liver failure, inflammatory bowel disease
 CC and Crohn's disease.

XX Sequence 176 AA;

Query Match 60.9%; Score 67; DB 21; Length 176;
 Best Local Similarity 61.1%; Pred. No. 0.42;

Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PATCANPSFVEGECPC 18
 I I I I I : I I I I

Db 150 patcavpnipgacpcvc 167

RESULT 14
AAM38630
ID AAM38630 standard; Protein; 270 AA.
XX
AC AAM38630;
XX
DT 19-OCT-2001 (first entry)
XX
XX Human colorectal cancer antigen SEQ ID NO: 145.
DE Human; colorectal cancer; colorectal cancer antigen; gene therapy.
XX Homo sapiens.
KW WO200155350-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01350.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 23-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.

05-DEC-2000; 2000US-0251030.
 05-DEC-2000; 2000US-0251988.
 05-DEC-2000; 2000US-0256719.
 06-DEC-2000; 2000US-0251479.
 08-DEC-2000; 2000US-0251856.
 08-DEC-2000; 2000US-0251868.
 08-DEC-2000; 2000US-0251869.
 08-DEC-2000; 2000US-0251989.
 08-DEC-2000; 2000US-0251990.
 11-DEC-2000; 2000US-0254097.
 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-457727/49.
 DR N-PSDB; AAI57608.
 XX Isolated polypeptide for treating, preventing and/or prognosing
 PT disorders related to the colon and rectum including colorectal cancers
 PT and also for testing and detection e.g. diagnosis -
 XX
 PS Claim 11; SEQ ID NO: 145; 522pp + Sequence Listing; English.
 CC The present invention provides the protein and coding sequences of a
 CC number of colorectal cancer antigens. These are shown in
 CC AAI57547-AAI57619 and AAI57619-AAI57641. These can be used in the
 CC diagnosis, prevention and treatment of cancer of the colon and/or rectum.
 CC The present sequence is a colorectal cancer antigen of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 270 AA;
 Query Match 60.9%; Score 67; DB 22; Length 270;
 Best Local Similarity 55.6%; Pred. No. 0.61;
 Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
 Qy 1 PATCANPSFVEGECPC 18
 Db 100 pttcarpgpggqcpc 117
 RESULT 15
 AAE03865
 ID AAE03865 standard; Protein; 550 AA.
 AC AAE03865;
 XX 08-AUG-2001 (first entry)
 XX Human gene 19 encoded secreted protein HOGCE48, SEQ ID NO:111.
 KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
 KW fetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
 KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
 KW inflammation; neurological disorder; Alzheimer's disease; food additive;
 KW angiotensin disorder; kidney disorder; gastrointestinal disorder; allergy;
 KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
 KW cell culture; chemotaxis; vulnery; binding partner identification;
 KW gene therapy; chromosome 2.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FT Peptide 1
 FT Protein 2..550

/note= "Mature secreted protein"
 WO200136440-A1.
 25-MAY-2001.
 15-NOV-2000; 2000WO-US31282.
 19-NOV-1999; 99US-0166414.
 21-JUL-2000; 2000US-0219665.
 (HUMA-) HUMAN GENOME SCI INC.
 PI Ruben SM, Komatsoulis GA, Birse CE, Moore PA;
 DR WPI; 2001-343795/36.
 DR N-PSDB; AAD08330.
 XX Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition -
 XX
 PS Claim 11; Page 523-525; 553pp; English.
 CC AAD08283-AAD08355 represent cDNAs corresponding to 23 human secreted
 CC protein genes, and AAE03818-AAE03870 represent the proteins they encode.
 CC AAE03871-AAE03896 represent human secreted protein fragments or variants.
 CC The secreted proteins and their genes are useful for preventing, gene
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 23 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein of the invention.
 XX Sequence 550 AA;

Query Match 60.9%; Score 67; DB 22; Length 550;
 Best Local Similarity 61.1%; Pred. No. 1.2;
 Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 Qy 1 PATCANPSFVEGECPC 18
 Db 524 patcavpvnipgaccpvc 541

Search completed: January 31, 2002, 09:23:05
 Job time: 195 sec

THIS PAGE BLANK (USPTO)

RESULT	2
Q9JLLO	
ID	Q9JLLO
AC	Q9JLLO; PRELIMINARY; PRT; 1028 AA.
DT	01-OCT-2000 (TREMBlrel. 15. Created)

RL DNA Res. 3:321-329(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RX MEDLINE=95048383; PubMed=7959781;
RA Weiler S.R., Taylor S.M., Deans R.J., Kan-Mitchell J., Mitchell M.S.,
RT Trent J.M.;
RA "Assignment of a human melanoma associated gene MG50 (D2S448) to
RT chromosome 2p25.3 by fluorescence in situ hybridization.";
RL Genomics 22:243-244(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RA Mitchell M.S., Kan-Mitchell J., Minev B., Edman C., Deans R.J.;
RT "Identification of a novel melanoma gene (MG50) - likely the gene for
RT IL-1 receptor antagonist - which encodes epitopes recognized by human
RT cytolytic T lymphocytes.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -|- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
DR EMBL; D86983; BAA13219.1; -;
DR EMBL; AF200348; AAF06354.1; -;
DR HSSP; P05164; 1CXP.
DR InterPro; IPR002007; Anim_peroxidase.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003591; LRR_Typ.
DR InterPro; IPR001536; Peroxidase_3.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF00047; ig; 4.
DR Pfam; PF00560; LRR; 5.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF00141; peroxidase; 1.
DR Pfam; PF00093; vwc; 1.
DR PRINTS; PR00457; ANPEROXIDASE.
DR SMART; SM00408; IGC2; 4.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_Typ; 4.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 1496 AA; 167209 MW; E9B9A7069BFIABFF CRC64;

Query Match 60.9%; Score 67; DB 4; Length 1496;
Best Local Similarity 61.1%; Pred. No. 0.067;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 PATCANPSFVEGCCPSC 18
Db 1470 PATCAVPVNI PGACCPVC 1487
||||| : |||||
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RX MEDLINE=95048383; PubMed=7959781;
RA Weiler S.R., Taylor S.M., Deans R.J., Kan-Mitchell J., Mitchell M.S.,
RT Trent J.M.;
RA "Assignment of a human melanoma associated gene MG50 (D2S448) to
RT chromosome 2p25.3 by fluorescence in situ hybridization.";
RL Genomics 22:243-244(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RA Mitchell M.S., Kan-Mitchell J., Minev B., Edman C., Deans R.J.;
RT "Identification of a novel melanoma gene (MG50) - likely the gene for
RT IL-1 receptor antagonist - which encodes epitopes recognized by human
RT cytolytic T lymphocytes.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -|- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
DR EMBL; D86983; BAA13219.1; -;
DR EMBL; AF200348; AAF06354.1; -;
DR HSSP; P05164; 1CXP.
DR InterPro; IPR002007; Anim_peroxidase.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003591; LRR_Typ.
DR InterPro; IPR001536; Peroxidase_3.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF00047; ig; 4.
DR Pfam; PF00560; LRR; 5.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF00141; peroxidase; 1.
DR Pfam; PF00093; vwc; 1.
DR PRINTS; PR00457; ANPEROXIDASE.
DR SMART; SM00408; IGC2; 4.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_Typ; 4.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 1496 AA; 167209 MW; E9B9A7069BFIABFF CRC64;

RC TISSUE=FLOOR PLATE;
RX MEDLINE=20266358; PubMed=10779551;
RA Matsui M., Mizuseki K., Nakatani J., Nakanishi S., Sasai Y.;
RT "Xenopus kielin: A dorsalizing factor containing multiple chordin-type
RT repeats secreted from the embryonic midline.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:5291-5296(2000).
DR EMBL; AB026192; BAA95483.1; -;
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002919; TIL.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001846; Vwd.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF01826; TIL; 1.
DR Pfam; PF00093; vwc; 22.
DR Pfam; PF00094; vwd; 1.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 27.
DR SMART; SM00011; VWC_dof; 1.
DR SMART; SM00216; VWD; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR PROSITE; PS01208; VWFC; 20.
SQ SEQUENCE 2327 AA; 255800 MW; 0293109329205983 CRC64;

Query Match 60.9%; Score 67; DB 13; Length 2327;
Best Local Similarity 50.0%; Pred. No. 0.098;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 PATCANPSFVEGCCPSC 18
Db 829 PSQSHPGKAAGQCCPSC 846
|:|:| |::|
RP SEQUENCE FROM N.A.
RC TISSUE=FLOOR PLATE;
RX MEDLINE=20266358; PubMed=10779551;
RA Matsui M., Mizuseki K., Nakatani J., Nakanishi S., Sasai Y.;
RT "Xenopus kielin: A dorsalizing factor containing multiple chordin-type
RT repeats secreted from the embryonic midline.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:5291-5296(2000).
DR EMBL; AB026192; BAA95483.1; -;
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002919; TIL.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001846; Vwd.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF01826; TIL; 1.
DR Pfam; PF00093; vwc; 22.
DR Pfam; PF00094; vwd; 1.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 27.
DR SMART; SM00011; VWC_dof; 1.
DR SMART; SM00216; VWD; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR PROSITE; PS01208; VWFC; 20.
SQ SEQUENCE 2327 AA; 255800 MW; 0293109329205983 CRC64;

RESULT 7
Q17429 PRELIMINARY; PRT; 909 AA.
AC Q17429;
DT 01-JAN-1999 (TrEMBLrel. 09, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE B0024.14 PROTEIN.
GN B0024.14.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z71178; CAA94886.2; -;
DR EMBL; AL021478; CAA94886.2; JOINED.
DR HSSP; P15358; ISKZ.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF00093; vwc; 4.
DR SMART; SM00214; VWC; 8.
DR PROSITE; PS01208; VWFC; UNKNOWN_1.
SQ SEQUENCE 909 AA; 99832 MW; 489CFCF4C39F1AA CRC64;

Query Match 60.0%; Score 66; DB 5; Length 909;
Best Local Similarity 66.7%; Pred. No. 0.062;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 CANPSFVEGCCPSC 18
Db 297 CENPRKVGCCPVC 311
| | | | | | | | | |
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RX MEDLINE=95048383; PubMed=7959781;
RA Weiler S.R., Taylor S.M., Deans R.J., Kan-Mitchell J., Mitchell M.S.,
RT Trent J.M.;
RA "Assignment of a human melanoma associated gene MG50 (D2S448) to
RT chromosome 2p25.3 by fluorescence in situ hybridization.";
RL Genomics 22:243-244(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RA Mitchell M.S., Kan-Mitchell J., Minev B., Edman C., Deans R.J.;
RT "Identification of a novel melanoma gene (MG50) - likely the gene for
RT IL-1 receptor antagonist - which encodes epitopes recognized by human
RT cytolytic T lymphocytes.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -|- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
DR EMBL; D86983; BAA13219.1; -;
DR EMBL; AF200348; AAF06354.1; -;
DR HSSP; P05164; 1CXP.
DR InterPro; IPR002007; Anim_peroxidase.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003591; LRR_Typ.
DR InterPro; IPR001536; Peroxidase_3.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF00047; ig; 4.
DR Pfam; PF00560; LRR; 5.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF00141; peroxidase; 1.
DR Pfam; PF00093; vwc; 1.
DR PRINTS; PR00457; ANPEROXIDASE.
DR SMART; SM00408; IGC2; 4.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_Typ; 4.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 1496 AA; 167209 MW; E9B9A7069BFIABFF CRC64;

DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE CHORDIN.
GN CHD.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Miller-Bertoglio V.E., Fisher S., Sanchez A., Mullins M.,
RA Halpern M.E.;
RL Dev. Biol. 0:0-0(1997).
DR EMBL: AF034606; AAB33485.1; -.
DR ZFIN: ZDB-GENE-990415-33; chd.
DR InterPro: IPR001007; VWFC.
DR Pfam: PF00093; vwc; 4.
DR PROSITE: PS01208; VWFC; 3.
DR SMART: SM00214; VWC; 4.
SQ SEQUENCE 940 AA; 104999 MW; B855CAF84F4623AC CRC64;

Query Match 57.3%; Score 63; DB 13; Length 940;
Best Local Similarity 56.2%; Pred. No. 0.18;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 TCANPSFVEGCCPSC 18
: | | | | | | | | | |
Db 903 SCRNPIRTGKCCPEC 918

RESULT 9
Q63079 PRELIMINARY; PRT; 1453 AA.
ID Q63079
AC Q63079
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE COLLAGEN ALPHA1 (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-1092 FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=TOOTH;
RA Brandsten C., Lundmark C., Christerson C., Hammarstrom L., Wurtz T.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z78279; CAB01633.1; -.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib_collagen_C.
DR InterPro: IPR001007; VWFC.
DR Pfam: PF01391; Collagen; 18.
DR Pfam: PF01410; COLFI; 1.
DR ProDom: PD02078; Fib_collagen_C; 1.
DR PROSITE: PS01208; VWFC; 1.
DR SMART: SM00038; COLFI; 1.
DR SMART: SM00214; VWC; 1.
FT NON_TER 1
SQ SEQUENCE 1453 AA; 137886 MW; E6896BDC19A4A1D8 CRC64;

Query Match 57.3%; Score 63; DB 11; Length 1453;
Best Local Similarity 56.7%; Pred. No. 0.26;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 CANPSFVEGCCPSC 18
: | | | | | | | | | |
Db 72 CPNFKREGCCPFC 86

RESULT 10
Q61431 PRELIMINARY; PRT; 1497 AA.
ID Q61431
AC Q61431
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE PROCOLLAGEN, TYPE V, ALPHA 2.
GN COL5A2 OR COL5A-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/C;
RX MEDLINE=93214071; PubMed=1297453;
RA Andrikopoulos K., Suzuki H.R., Solursh M., Ramirez F.;
RT "Localization of pro-alpha 2(V) collagen transcripts in the tissues of
the developing mouse embryo";
RL Dev. Dyn. 195:113-120(1992).
DR EMBL: L02918; AAA37440.1; -.
DR MGD; MGI:88458; Col5a2.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib_collagen_C.
DR InterPro: IPR001007; VWFC.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 18.
DR Pfam: PF00093; vwc; 1.
DR ProDom: PD02078; Fib_collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
DR SMART: SM00214; VWC; 1.
DR PROSITE: PS01208; VWFC; 1.
SQ SEQUENCE 1497 AA; 144966 MW; 343D30B6C85F22A3 CRC64;

Query Match 56.4%; Score 62; DB 11; Length 1497;
Best Local Similarity 66.7%; Pred. No. 0.38;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 CANPSFVEGCCPSC 18
: | | | | | | | | | |
Db 81 CANPITPPGCCPVC 95

RESULT 11
Q9VPD1 PRELIMINARY; PRT; 434 AA.
ID Q9VPD1
AC Q9VPD1
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE CG4786 PROTEIN.
GN CG4786.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20195006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merklov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003591; AAF51624.1; -.
 DR FlyBase; FBgn0037012; CG4786.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001007; VWFC.
 DR Pfam; PF00093; vwc; 1.
 DR SMART; SM00214; vwc; 1.
 DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
 SQ SEQUENCE 434 AA; 47737 MW; 8D233AD5F611F017 CRC64;

Query Match 54.5%; Score 60; DB 5; Length 434;
 Best Local Similarity 60.0%; Pred. No. 0.26;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 CANPFSFVEGECPC 18
 DB 136 CNNPVEVKGCCPVC 150
 ID 093251 PRELIMINARY; PRT; 1445 AA.
 AC 093251;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE ALPHA 1 TYPE I COLLAGEN.
 OS Rana catesbeiana (Bull frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Rana.
 OX NCBI_TaxID=8400;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Asahina K., Utoh R., Obara M., Yoshizato K.;
 RA "Spatial-temporal expression of bullfrog $\alpha 1(I)$ and $\alpha 2(I)$ collagen genes
 RT in intestine during metamorphosis.";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB015440; BAA29028.1; -.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF01410; COLFI; 1.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR PROSITE; PS01208; VWFC; 1.

DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 SQ SEQUENCE 1445 AA; 137251 MW; F59BB550C9873F04 CRC64;

Query Match 54.5%; Score 60; DB 13; Length 1445;
 Best Local Similarity 60.0%; Pred. No. 0.72;
 Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 CANPFSFVEGECPC 18
 DB 74 CPNPEIPMGCCPVC 88
 ID 090793 PRELIMINARY; PRT; 67 AA.
 AC 090793;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE COLLAGEN (FRAGMENT).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84270696; PubMed=6547770;
 RA Yamada Y., Liaw G., Mudryj M., Obici S., de Crombrughe B.;
 RT "Conservation of the sizes for one but not another class of exons in
 RL two chick collagen genes.";
 DR EMBL; X00821; CAA25395.1; -.
 DR Nature 310:333-337(1984).
 DR InterPro; IPR001007; VWFC.
 DR Pfam; PF00093; vwc; 1.
 DR PROSITE; PS01208; VWFC; 1.
 DR SMART; SM00214; VWC; 1.
 FT NON_TER 1 67
 SQ SEQUENCE 67 AA; 7323 MW; 484BB20152BEB61 CRC64;

Query Match 53.6%; Score 59; DB 13; Length 67;
 Best Local Similarity 60.0%; Pred. No. 0.075;
 Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 CANPFSFVEGECPC 18
 DB 44 CPNPEIPMGCCPVC 58
 ID 090802 PRELIMINARY; PRT; 206 AA.
 AC 090802;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE COLLAGEN TYPE 2 PRECURSOR (FRAGMENT).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92078225; PubMed=1744138;
 RA Nah H.D., Upholt W.B.;
 RT "Type II collagen mRNA containing an alternatively spliced exon
 RL predominates in the chick limb prior to chondrogenesis.";
 RL J. Biol. Chem. 266:23446-23452(1991).

Search completed: January 31, 2002, 09:36:15
Job time: 784 sec

DR EMBL; M74435; AAA48714.1; -.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00093; vwc; 1.
DR Pfam; PF01391; Collagen; 1.
DR PROSITE; PS01208; VWFC; 1.
DR SMART; SM00214; VWC; 1.
KW Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 185 >206 COLLAGEN TYPE 2.
FT NON_TER 206 206
SQ SEQUENCE 206 AA; 20793 MW; E3EB84ACF0C31459 CRC64;

Query Match 53.6%; Score 59; DB 13; Length 206;
Best Local Similarity 50.0%; Pred. No. 0.19; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 PATCANPSFVEGCCPSC 18
| | | | |
DB 75 PDCPSPEIPFGCCPVC 92

RESULT 15
Q15667
ID Q15667 PRELIMINARY; PRT; 376 AA.
AC Q15667;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE THROMBOSPONDIN (TSP) PRECURSOR (FRAGMENT).
GN THBS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86287276; PubMed=3461443;
RA Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
RT "Characterization of a cDNA encoding the heparin and collagen binding
domains of human thrombospondin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92278426; PubMed=1350660;
RA Burglin T.R., Barnes T.N.;
RT "Introns in sequence tags.";
RL Nature 357:367-368(1992).
DR EMBL; M14326; AAA61237.1; -.
DR InterPro; IPR001007; VWFC.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF00093; vwc; 1.
DR Pfam; PF02210; TSPN; 1.
DR PROSITE; PS01208; VWFC; 1.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 >376 POTENTIAL.
FT NON_TER 376 376
SQ SEQUENCE 376 AA; 41329 MW; 313B62C0D253BA2B CRC64;

Query Match 53.6%; Score 59; DB 4; Length 376;
Best Local Similarity 53.3%; Pred. No. 0.33;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 CANPSFVEGCCPSC 18
| | | | |
DB 358 CSNATVPDGCPCRC 372

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:23:55 ; Search time 43.81 Seconds
(without alignments)
9.246 Million cell updates/sec

Title: US-09-536-087-7

Perfect score: 110

Sequence: 1 PATCANPSFVEGECPCSC 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/6C_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/6D_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	95.5	1172	1	US-08-313-288B-19
2	67	60.9	176	2	US-08-726-525-2
3	67	60.9	176	2	US-08-487-942-2
4	67	60.9	176	2	US-08-726-036A-2
5	67	60.9	176	4	US-09-083-516-2
6	63.5	57.7	810	2	US-08-820-170A-34
7	63.5	57.7	810	3	US-09-055-699-34
8	63.5	57.7	810	4	US-09-273-565-34
9	59	53.6	218	3	US-08-985-526-1
10	59	53.6	239	5	PCT-US93-01652-1
11	59	53.6	441	3	US-08-985-526-3
12	59	53.6	1170	1	US-08-313-288B-20
13	59	53.6	1442	2	US-08-316-650-12
14	59	53.6	1442	5	PCT-US95-02251-12
15	53	48.2	941	1	US-08-938-365-4
16	53	48.2	941	1	US-08-343-760A-2
17	52	47.3	160	1	US-08-479-233-11
18	52	47.3	160	5	PCT-US93-00643-11
19	52	47.3	1341	3	US-08-963-825-18
20	51	46.4	623	4	US-09-029-348-3
21	51	46.4	626	4	US-09-029-348-2
22	50	45.5	816	2	US-08-820-170A-37
23	50	45.5	816	3	US-09-055-699-37
24	50	45.5	816	4	US-09-273-565-37
25	50	45.5	855	2	US-08-938-365-3
26	50	45.5	867	2	US-08-938-365-2
27	50	45.5	954	2	US-08-749-169A-3

28	50	45.5	954	2	US-09-130-032A-3	Sequence 3, Appli
29	47.5	43.2	194	1	US-08-616-368A-11	Sequence 11, Appl
30	47.5	43.2	194	4	US-09-054-298-11	Sequence 11, Appl
31	47.5	43.2	194	4	US-08-818-655-11	Sequence 11, Appl
32	46	41.8	409	1	US-08-331-394-2	Sequence 2, Appli
33	46	41.8	409	1	US-08-250-858-2	Sequence 2, Appli
34	46	41.8	409	1	US-08-446-915-2	Sequence 2, Appli
35	46	41.8	409	2	US-08-744-139-2	Sequence 2, Appli
36	46	41.8	409	5	PCT-US95-06639-2	Sequence 2, Appli
37	45.5	41.4	651	1	US-08-264-101-2	Sequence 2, Appli
38	45.5	41.4	651	2	US-08-765-243-2	Sequence 2, Appli
39	45.5	41.4	651	5	PCT-US95-07295-2	Sequence 2, Appli
40	45.5	41.4	734	2	US-08-765-243-8	Sequence 8, Appli
41	45.5	41.4	734	5	PCT-US95-07295-8	Sequence 8, Appli
42	44	40.0	316	4	US-09-111-470-4	Sequence 4, Appli
43	43.5	39.5	194	2	US-08-739-485-8	Sequence 8, Appli
44	43	39.1	846	1	US-08-149-103-3	Sequence 3, Appli
45	43	39.1	846	1	US-08-451-883-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-313-288B-19
; Sequence 19, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; NUMBER OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1172 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-313-288B-19

Query Match 95.5%; Score 105; DB 1; Length 1172;
Best Local Similarity 94.4%; Pred. No. 3.2e-05;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 PATCANPSFVEGECPCSC 18
Db 357 PATCANPSFVEGECPCSC 374

RESULT 2
US-08-726-525-2
; Sequence 2, Application US/08726525
; Patent No. 5789181
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; TITLE OF INVENTION: BINDING
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726.525
; FILING DATE: 07-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,942
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15258
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-726-525-2

Query Match 60.9%; Score 67; DB 1; Length 176;
Best Local Similarity 61.1%; Pred. No. 0.19;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 1 PATCANPSFVEGECPC 18
||||| : |||||
DB 150 PATCAVPVNPICACCPVC 167
RESULT 3
US-08-487-942-2
; Sequence 2, Application US/08487942
; Patent No. 5817476
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; TITLE OF INVENTION: BINDING
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA

COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,942
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15258
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 176 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-487-942-2
Query Match 60.9%; Score 67; DB 2; Length 176;
Best Local Similarity 61.1%; Pred. No. 0.19;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 1 PATCANPSFVEGECPC 18
||||| : |||||
DB 150 PATCAVPVNPICACCPVC 167
RESULT 4
US-08-726-036A-2
; Sequence 2, Application US/08726036A
; Patent No. 5981482
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; TITLE OF INVENTION: BINDING
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726.036A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15258
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-726-036A-2

```
;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-726-036A-2

Query Match 60.9%; Score 67; DB 2; Length 176;
Best Local Similarity 61.1%; Pred. No. 0.19;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 PATCANPSFVEGCCPC 18
| | | | | : | | | |
Db 150 PATCAVPVNI PGACPC 167

RESULT 5
US-09-083-516-2
; Sequence 2, Application US/09083516
; Patent No. 6300086
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; TITLE OF INVENTION: BINDING
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,516
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,942
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15258
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-083-516-2

Query Match 60.9%; Score 67; DB 4; Length 176;
Best Local Similarity 61.1%; Pred. No. 0.19;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 PATCANPSFVEGCCPC 18
| | | | | : | | | |
Db 150 PATCAVPVNI PGACPC 167

RESULT 6
US-08-820-170A-34
; Sequence 34, Application US/08820170A
```

```
; Patent No. 5831058
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,170A
; FILING DATE:
; CLASSIFICATION: 536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 810 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-820-170A-34

Query Match 57.7%; Score 63.5; DB 2; Length 810;
Best Local Similarity 52.2%; Pred. No. 2;
Matches 12; Conservative 1; Mismatches 5; Indels 5; Gaps 1;

QY 1 PATCANPS-----FVEGCCPC 18
| | | | | : | | | |
Db 727 PLTCPNLSCYTAILEGCCPC 749

RESULT 7
US-09-055-699-34
; Sequence 34, Application US/09055699
; Patent No. 6005088
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,699
; FILING DATE:
; CLASSIFICATION:
```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/820,170
; FILING DATE:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 810 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-055-699-34

Query Match 57.7%; Score 63.5; DB 3; Length 810;
Best Local Similarity 52.2%; Pred. No. 2;
Matches 12; Conservative 1; Mismatches 5; Indels 5; Gaps 1;

Qy 1 PATCANPS-----FVEGECPCSC 18
Db 727 PLTCPNLSCYEYTAILEGECPCRC 749

RESULT 8
US-09-273-565-34
; Sequence 34, Application US/09273565A
; Patent No. 6166190
; GENERAL INFORMATION:
; APPLICANT: FUJIWARA, TSUROMU
; APPLICANT: WATANABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
; FILE REFERENCE: Q-53599
; CURRENT APPLICATION NUMBER: US/09/273,565A
; CURRENT FILING DATE: 1999-03-22
; EARLIER APPLICATION NUMBER: 09/055,699
; EARLIER FILING DATE: 1998-04-07
; EARLIER APPLICATION NUMBER: 08/820,170
; EARLIER FILING DATE: 1997-03-19
; EARLIER APPLICATION NUMBER: JP 63410/1996
; EARLIER FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: JP 69163/1997
; EARLIER FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 34
; LENGTH: 810
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-273-565-34

Query Match 57.7%; Score 63.5; DB 4; Length 810;
Best Local Similarity 52.2%; Pred. No. 2;
Matches 12; Conservative 1; Mismatches 5; Indels 5; Gaps 1;

Qy 1 PATCANPS-----FVEGECPCSC 18
Db 727 PLTCPNLSCYEYTAILEGECPCRC 749

RESULT 9
US-08-985-526-1
; Sequence 1, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER-DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY

; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorow Jr., Robert G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-985-526-1

Query Match 53.6%; Score 59; DB 3; Length 218;
Best Local Similarity 53.3%; Pred. No. 2.1;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 4 CANPSFVEGECPCSC 18
Db 58 CSNATVPDGECCPCRC 72

RESULT 10
PCT-US93-01652-1
; Sequence 1, Application PC/TUS9301652
; GENERAL INFORMATION:
; APPLICANT: Bouck, Noel P.
; APPLICANT: Polverini, Peter J.
; APPLICANT: Good, Deborah J.
; APPLICANT: Frazier, William A.
; TITLE OF INVENTION: Method and Composition for
; TITLE OF INVENTION: Inhibiting Angiogenesis
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut
; STREET: 100 South Wacker Drive, Suite 960
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-4002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/01652
; FILING DATE: 19930222
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/841,656
; FILING DATE: 24-FEB-1992
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/464,369
FILING DATE: 12-JAN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Featress, Susan B.
REGISTRATION NUMBER: 31,327
REFERENCE/DOCKET NUMBER: 92005-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)-456-8000
TELEFAX: (312)-456-7776
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 239 amino acids
TYPE: AMINO ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
PCT-US93-01652-1

Query Match 53.6%; Score 59; DB 5; Length 239;
Best Local Similarity 53.3%; Pred. No. 2.3;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 CANPSFVEGCCPSC 18
| : : ||||| |
Db 47 CSNATVPDGECCPRC 61

RESULT 11
US-08-985-526-3
Sequence 3, Application US/08985526
Patent No. 6080728
GENERAL INFORMATION:
APPLICANT: Mixson, James A.
TITLE OF INVENTION: CARRIER-DNA COMPLEXES CONTAINING DNA
TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
TITLE OF INVENTION: THERAPY
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Connolly, Bove, Lodge, & Hutz
STREET: 1220 Market Street, P.O. Box 2207
CITY: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
ZIP: 19899
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,526
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,845
FILING DATE: 16-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: McMorro Jr., Robert G.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 658-9141
TELEFAX: (302) 658-5613
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 441 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-985-526-3

Query Match 53.6%; Score 59; DB 3; Length 441;
Best Local Similarity 53.3%; Pred. No. 3.9;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 CANPSFVEGCCPSC 18
| : : ||||| |
Db 58 CSNATVPDGECCPRC 72

RESULT 12
US-08-313-288B-20
Sequence 20, Application US/08313288B
Patent No. 5750502
GENERAL INFORMATION:
APPLICANT: Jessell, Thomas M. and Avi Hu Klar
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/313,288B
FILING DATE: January 5, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0526
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1170 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-313-288B-20

Query Match 53.6%; Score 59; DB 1; Length 1170;
Best Local Similarity 53.3%; Pred. No. 9.4;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 CANPSFVEGCCPSC 18
| : : ||||| |
Db 358 CSNATVPDGECCPRC 372

RESULT 13
US-08-316-650-12
Sequence 12, Application US/08316650
Patent No. 5942496
GENERAL INFORMATION:
APPLICANT: Bonadio, Jeffrey
APPLICANT: Roessler, Blake J.
APPLICANT: Goldstein, Steven A.
APPLICANT: Lin, Wushan
TITLE OF INVENTION: METHODS AND COMPOSITIONS
TITLE OF INVENTION: FOR STIMULATING BONE CELLS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433

;
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/316,650
; FILING DATE: 30-SEP-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/199,780
; FILING DATE: 30-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UMIC:008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1442 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US95-02251-12

Query Match 53.6%; Score 59; DB 2; Length 1442;

Best Local Similarity 60.0%; Pred. No. 11;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 4 CANPSFVEGCCPC 18
| | | | | | | | | |
DB 74 CLNPEIPFGCCPIC 88

RESULT 14
PCT-US95-02251-12
; Sequence 12, Application PC/TUS9502251
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE
; TITLE OF INVENTION: CELLS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02251
; FILING DATE: CONCURRENTLY HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/316,650
; FILING DATE: 30-SEP-1994
; CLASSIFICATION:
; APPLICATION NUMBER: US 08/199,780

;
; FILING DATE: 18-FEB-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UMIC009P--
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1442 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US95-02251-12

Query Match 53.6%; Score 59; DB 5; Length 1442;

Best Local Similarity 60.0%; Pred. No. 11;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 4 CANPSFVEGCCPC 18
| | | | | | | | | |
DB 74 CLNPEIPFGCCPIC 88

RESULT 15
US-08-938-365-4
; Sequence 4, Application US/08938365
; Patent No. 5989909
; GENERAL INFORMATION:
; APPLICANT: Yang, Pan
; TITLE OF INVENTION: HUCHORDIN AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,365
; FILING DATE: 26-SEP-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 09404/040001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 940 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-938-365-4

Query Match 48.2%; Score 53; DB 2; Length 940;
Best Local Similarity 38.9%; Pred. No. 40;
Matches 7; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 PATCANPSFVEGCCPSC 18
| : | : : | | | |
Db 732 PLNCSPVHLFDQCCPVC 749

Search completed: January 31, 2002, 09:23:56
Job time: 221 sec

THIS PAGE BLANK (USPTO)

SUMMARIES

PT activity -
 XX Disclosure; Page 40; 73pp; English.
 PS
 CC The present sequence is that of peptide 3, derived from the
 CC procollagen domain (amino acids 327-348) of human thrombospondin-2
 CC (TSP-2, see AAB19677). The peptide is 1 of 5 synthetic peptides (see
 CC AAB19679-83) used with human dermal microvascular endothelial cells
 CC (HMEC) to determine the effect of TSP-2 on endothelial cell
 CC migration. Peptide 3 did not significantly modify HMEC migration,
 CC in contrast to peptide 7 (see AAB19683) from a type 1 repeat of
 CC TSP-2, which inhibited migration. The invention is based on the
 CC discovery that overexpression of TSP-2 decreases tumor size in
 CC vivo, and features methods for modulating unwanted angiogenesis and
 CC cell proliferation by increasing TSP-2 activity.
 XX Sequence 22 AA;
 SQ

Query Match 100.0%; Score 128; DB 21; Length 22;
 Best Local Similarity 100.0%; Pred. No. 4.9e-10;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FAENETWVDSCTTCTCKKFKT 22
 DB 1 faenetwvdscttctckkfk 22
 |||||

RESULT 2
 AAB19677
 ID AAB19677 standard; Protein: 1172 AA.
 XX
 AC AAB19677;
 DT 05-FEB-2001 (first entry)
 XX
 DE Human thrombospondin-2.
 XX
 KW Thrombospondin-2; TSP-2; human; angiogenesis; cell proliferation;
 KW melanoma; tumor; cancer; squamous cell carcinoma; antiangiogenic;
 KW prostate cancer; psoriasis; rosacea dermatosis; antitumor;
 KW therapy.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 382..429
 FT /note= "type 1 repeat"
 FT Region 384..390
 FT /note= "WSPWAEW sequence involved in antiangiogenic
 FT activity"
 FT Region 438..490
 FT /note= "type 1 repeat"
 FT Region 495..547
 FT /note= "type 1 repeat"
 XX
 PN WO200057899-A1.
 XX
 PD 05-OCT-2000.
 XX
 PF 24-MAR-2000; 2000WO-US07835.
 XX
 PR 31-MAR-1999; 99US-0127221.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Detmar M, Streit M;
 XX
 DR WPI; 2000-656131/63.
 DR N-PSDB; AAA88669.
 XX
 PT Treating a disorder characterized by unwanted cell proliferation e.g.
 PT precancerous, cancerous or neoplastic cells or presence of tumor

PT preferably of skin or prostate, comprises increasing thrombospondin-2
 PT activity -
 XX Disclosure; Fig 2; 73pp; English.
 PS
 CC The present sequence is that of human thrombospondin-2 (TSP-2).
 CC The invention is based on the discovery that overexpression of
 CC TSP-2 decreases tumor size in vivo, and features methods for
 CC modulating unwanted angiogenesis and tumor growth. Treatment of
 CC unwanted cell proliferation or angiogenesis involves increasing
 CC TSP-2 activity. This is achieved by administering an agent which
 CC increases TSP-2 activity, especially a TSP-2 polypeptide, a TSP-2
 CC derived polypeptide or retro-inverso peptide, a nucleic acid
 CC encoding TSP-2, an agonist of TSP-2, or an agent that increases
 CC TSP-2 gene expression. The TSP-2 polypeptide may include at least
 CC one type 1 repeat such as the WSPWAEW peptide (see AAB19683),
 CC which is involved in the antiangiogenic activity of TSP-2. The
 CC method is used to treat a disorder characterised by pre-cancerous,
 CC cancerous or neoplastic cells, or the presence of a tumour, or a
 CC disorder that affects epithelial tissues resulting in unwanted
 CC skin cell proliferation. Such disorders include malignant
 CC melanoma, prostate cancer, squamous cell carcinoma, aged skin,
 CC rosacea dermatosis, psoriasis, and skin damage caused by
 CC photoradiation (all claimed). Evaluating the presence of TSP-2
 CC nucleic acid or protein is useful for diagnosing a subject at risk
 CC of unwanted cell proliferation or angiogenesis. Methods are also
 CC provided of identifying compounds that modulate TSP-2 activity.
 XX Sequence 1172 AA;
 SQ

Query Match 100.0%; Score 128; DB 21; Length 1172;
 Best Local Similarity 100.0%; Pred. No. 1.9e-08;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FAENETWVDSCTTCTCKKFKT 22
 DB 327 faenetwvdscttctckkfk 348
 |||||

RESULT 3
 AAB00043
 ID AAB00043 standard; Protein: 1172 AA.
 XX
 AC AAB00043;
 DT 08-NOV-2000 (first entry)
 XX
 DE Human thrombospondon-2 (TSP-2).
 XX
 KW TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
 KW thrombospondin; angiogenesis; tumour; treatment; cancer;
 KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
 KW glaucoma.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 381..436
 FT /label= Type 1 repeat region
 FT Region 437..493
 FT /label= Type 1 repeat region
 FT Region 494..550
 FT /label= Type 1 repeat region
 XX
 PN WO200044908-A2.
 XX
 PD 03-AUG-2000.
 XX
 PF 01-FEB-2000; 2000WO-US02482.
 XX
 PR 01-FEB-1999; 99US-0118053.
 XX

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 XX Lawler JW;
 PI WPI; 2000-514823/46.
 XX
 DR Nucleic acids encoding chimeric proteins such as cartilage oligomeric
 XX matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for
 PT inhibiting angiogenesis and treating diseases such as cancer
 PT
 PS Disclosure; Fig 2; 40pp; English.
 XX
 CC New nucleic acids are described which encode a protein comprising
 CC the second and third type 1 repeats of human TSP (thrombospondin)-1,
 CC but not the TGF (transforming growth factor)-beta activation region
 CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing
 CC the second and third type-1 repeats and the COMP (cartilage
 CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was
 CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1
 CC caused inhibition of the growth of tumours in mice models.
 CC Thus the nucleic acids and proteins may be useful for treating
 CC angiogenesis related diseases such as cancer (by reducing the rate of
 CC growth and size of tumours), arthritis, psoriasis, diabetic
 CC retinopathy, corneal graft rejection, and glaucoma. They may also be
 CC used for treating human immunodeficiency virus (HIV) infection.
 CC Anti-angiogenic therapy has little toxicity, does not require the
 CC therapeutic agent to enter tumour cells or cross the blood-brain
 CC barrier, controls tumour growth independently of growth of
 CC tumour cell heterogeneity, and does not induce drug resistance.
 XX
 SQ Sequence 1172 AA;

Query Match 100.0%; Score 128; DB 21; Length 1172;
 Best Local Similarity 100.0%; Pred. No. 1.9e-08;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FAENETWVDSCTTCTCKKFKT 22
 DB 327 faenetwvdscttctckkfkf 348

RESULT 4
 AAM15926
 ID AAM15926 standard; Protein; 63 AA.
 XX
 AC AAM15926;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Peptide #2360 encoded by probe for measuring cervical gene expression.
 XX
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 XX
 OS Homo sapiens.
 XX
 XX WO200157278-A2.
 PN
 XX 09-AUG-2001.
 PD
 XX 30-JAN-2001; 2001WO-US00670.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234887.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-48897/53.
 DR
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488901/53.
 XX
 DR Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 PT
 PS Claim 27; SEQ ID NO 20752; 487pp; English.
 XX
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs: see AAI10068-AAI28459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human Hela cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 63 AA;

Query Match 60.9%; Score 78; DB 22; Length 63;
 Best Local Similarity 70.6%; Pred. No. 0.0031;
 Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 FAENETWVDSCTTCTC 17
 DB 15 fvegetwnidscqtctc 31

RESULT 5
 AAM28429
 ID AAM28429 standard; Protein; 63 AA.
 XX
 AC AAM28429;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Peptide #2466 encoded by probe for measuring placental gene expression.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-48897/53.
 DR
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 PT
 PS Claim 27; SEQ ID NO 28698; 654pp; English.
 XX

224

CC undifferentiated cell populations, such as embryonic cells, or stem
 CC cells, to enhance growth and/or differentiation of the cells.
 CC The proteins may also have other useful properties characteristic of
 CC the TGF-beta superfamily of proteins. Such properties include
 CC angiogenic, chemotactic, and/or chemoattractant properties, and
 CC effects on cells including induction or inhibition of collagen
 CC synthesis, fibrosis, differentiation responses, cell proliferative
 CC responses, and responses involving cell adhesion, migration, and
 CC extracellular matrices. These properties make the proteins potential
 CC agents for wound healing, reduction of fibrosis, and reduction of
 CC scar tissue formation. Chordin-related proteins may also be useful
 CC for advancement of the onset of fertility in sexually immature
 CC mammals, so as to increase the lifetime reproductive performance of
 CC domestic animals such as cows, sheep and pigs. Chordin-related
 CC proteins may also be useful in modulating hematopoiesis by inducing
 CC the differentiation of erythroid cells, for suppressing the
 CC development of gonadal tumors, or for augmenting the activity of
 CC BMPs. The proteins may also have value as a dietary supplement, or
 CC as a component of cell culture media.
 XX
 SQ Sequence 400 AA;

Query Match 60.9%; Score 78; DB 21; Length 400;
 Best Local Similarity 70.6%; Pred. No. 0.017;
 Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 FAENETWVDSCTTC 17
 | | | | : | | | | |
 Db 51 fvegetwnidscgtc 67

RESULT 8
 AAY53033
 ID AAY53033 standard; Protein; 400 AA.
 AC AAY53033;
 XX
 XX 29-FEB-2000 (first entry)
 DT
 DE Human secreted protein clone dj167_2 protein sequence SEQ ID NO:72.
 XX Human; secreted protein; nutritional; cytokine; cell proliferation;
 KW differentiation; immune stimulating; vaccine; suppression;
 KW hematopoiesis regulation; tissue growth; activin; inhibin;
 KW chemotactic; chemokinetic; haemostatic; thrombolytic; receptor;
 KW ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
 KW tumour inhibition; gene therapy.
 XX Homo sapiens.
 OS
 XX WO957132-A1.
 PN
 XX 11-NOV-1999.
 PD
 XX 07-MAY-1999; 99WO-US09970.
 PF
 XX 07-MAY-1998; 98US-0084564.
 PR 02-JUN-1998; 98US-0087645.
 PR 22-JUL-1998; 98US-0093712.
 PR 31-JUL-1998; 98US-0094935.
 PR 11-AUG-1998; 98US-0095880.
 PR 11-AUG-1998; 98US-0096068.
 PR 06-MAY-1999; 99US-0096068.
 XX
 XX (GEM) GENETICS INST INC.
 PA
 XX Jacobs K, McCoy JM, LaVallie ER, Collins-Pacie LA, Evans C;
 PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Bowman MR;
 PI DiBasio-Smith E, Widom A;
 XX WPI; 2000-052937/04.
 DR N-PSDB; AAZ33351.
 DR

XX New polynucleotides encoding secreted human proteins, derived from
 PT adult placenta, adult retina, fetal brain, fetal
 XX Claim 81; Page 423-424; 492pp; English.
 XX The present invention describes new human secreted proteins which were
 CC isolated from adult placenta, adult retina, foetal brain, foetal kidney,
 CC adult blood, adult brain, adult thyroid, adult bladder, adult neural
 CC tissue, adult testes, and adult lymph node cDNA libraries. The human
 CC secreted proteins, and the polynucleotides encoding them, are predicted
 CC to have biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals. Suggested activities include nutritional activity, cytokine
 CC and cell proliferation/differentiation activity, immune stimulating
 CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
 CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
 CC activity, receptor/ligand activity, anti-inflammatory activity,
 CC cadherin/tumour invasion suppressor activity, and tumour inhibition
 CC activity. The polynucleotides are also stated to be useful for gene
 CC therapy. AAZ33316 to AAZ33373 encode human secreted proteins, and
 CC AAY52998 to AAY53060 represent human secreted proteins, given in the
 CC present invention.
 XX
 SQ Sequence 400 AA;

Query Match 60.9%; Score 78; DB 21; Length 400;
 Best Local Similarity 70.6%; Pred. No. 0.017;
 Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 FAENETWVDSCTTC 17
 | | | | : | | | | |
 Db 51 fvegetwnidscgtc 67

RESULT 9
 AAB61140
 ID AAB61140 standard; Protein; 732 AA.
 XX
 AC AAB61140;
 XX
 XX 30-MAR-2001 (first entry)
 DT
 XX Human NOV10 protein.
 DE
 XX Human; NOVX; antiinflammatory; cytostatic; neuroprotective;
 KW cerebroprotective; immunomodulator; vulnery; vasotropic; gene therapy;
 KW hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;
 KW diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.
 XX Homo sapiens.
 OS
 XX WO200075321-A2.
 PN
 XX 14-DEC-2000.
 PD
 XX 01-JUN-2000; 2000WO-US15303.
 PF
 XX 03-JUN-1999; 99US-0137322.
 PR 16-MAR-2000; 2000US-0189810.
 PR 22-MAR-2000; 2000US-0191158.
 PR 30-MAR-2000; 2000US-0193086.
 PR 31-MAY-2000; 2000US-0137322.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 XX Shimkets RA, Fernandes E, Herrman J, Vernet C;
 PI WPI; 2001-102403/11.
 XX N-PSDB; AAF27858.
 DR
 XX

PT New NOVX polypeptides and polynucleotides, useful in gene therapy, as a
 PT diagnostic marker, protein therapeutic, antibody or small molecule drug
 PT target for treating immune, proliferative and metabolic diseases and
 PT wound healing
 XX
 PS Claim 1; Page 39-42; 194pp; English.
 XX
 CC The present sequence is a new isolated polypeptide (NOVX). The NOVX
 CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
 CC treating or preventing NOVX-associated disorders. They are also useful
 CC for determining the presence of or a predisposition to a disease
 CC associated with altered levels of the NOVX polypeptide or nucleic acid.
 CC These NOVX-associated disorders include hyperplasias, tumours,
 CC resenosis, psoriasis, Dupuytren's contracture, diabetic complications,
 CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
 CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
 CC are especially useful in gene therapy. Specifically, NOVX is useful as
 CC a diagnostic marker or prognostic marker, protein therapeutic and
 CC antibody target or small molecule drug target to treat disorders in the
 CC immune response pathway, thyroid and metabolic diseases, bone metabolic
 CC disorders, diseases of the pancreas (e.g. diabetes or digestive
 CC disorders), proliferative diseases, or tissue regeneration and
 CC development (e.g. wound healing or treatment of burns).
 XX
 SQ Sequence 732 AA;
 Query Match 60.9%; Score 78; DB 22; Length 732;
 Best Local Similarity 70.6%; Pred. No. 0.03;
 Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 FAENETWVVDSTCTC 17
 Db 529 fvegetwnidstctc 545
 RESULT 10
 ID AAY82776 standard; Protein; 1036 AA.
 AC AAY82776;
 XX
 DT 19-JUN-2000 (first entry)
 XX
 DE Human chordin related protein (Clone dj167_19).
 KW Chordin related protein; cartilage; bone; connective tissue;
 KW periodontal disease; osteoporosis; burn; incision; ulcer; neuron;
 KW astrocyte; glial cell; transplantation; nerve; epidermis; muscle;
 KW liver; brain; lung; cardiac; pancreas; kidney; growth;
 KW differentiation; TGF-Beta; angiogenesis; chemotaxis;
 KW chemoattraction; collagen synthesis; fibrosis; cell adhesion;
 KW cell migration; fertility; reproduction; hematopoiesis;
 KW erythroid cell; tumour; dietary supplement; growth medium.
 XX
 OS Homo sapiens.
 XX
 PN WO200009551-A1.
 XX
 PD 24-FEB-2000.
 XX
 PF 10-AUG-1999; 99WO-US18117.
 XX
 PR 10-AUG-1998; 98US-0095880.
 PR 06-MAY-1999; 99US-0306111.
 XX
 PA (GEMV) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-racie LA, Merberg D;
 PI Treacy M, Diblasio-smith E, Widom A;
 XX
 DR WPI: 2000-205978/18.
 DTC N-PSDB; AA293172.

XX New polynucleotides encoding secreted human proteins, useful for
 PT treating e.g. broken bones, craniofacial defects, periodontal disease,
 PT osteoporosis, burns, incisions or ulcers
 XX
 PS Claim 21; Page 94-98; 105pp; English.
 XX
 CC The human chordin related protein and polynucleotides encoding them
 CC are predicted to have biological activities which would make them
 CC suitable for treating, preventing or ameliorating medical conditions
 CC which involve defects in cartilage, bone or connective tissue
 CC formation and damage to cartilage, bone or connective tissue, e.g.
 CC broken bones, congenital, trauma-induced, or
 CC oncologic-resection-induced craniofacial defects, periodontal
 CC disease, defects in the periodontal ligament or attachment apparatus,
 CC damage to the periodontal ligament or attachment apparatus,
 CC osteoporosis, burns, incisions or ulcers. The proteins may also
 CC affect neuronal, astrocytic, and glial cell survival and therefore be
 CC useful in transplantation and treatment of conditions exhibiting a
 CC decrease in neuronal survival and repair. The proteins may also be
 CC useful for the treatment of conditions related to other types of
 CC tissue, such as nerve, epidermis, muscle, and other organs such as
 CC liver, brain, lung, cardiac, pancreas, and kidney tissue. The
 CC proteins may further be useful for the treatment of relatively
 CC undifferentiated cell populations, such as embryonic cells, or stem
 CC cells, to enhance growth and/or differentiation of the cells.
 CC The proteins may also have other useful properties characteristic of
 CC the TGF-beta superfamily of proteins. Such properties include
 CC angiogenic, chemotactic, and/or chemoattractant properties, and
 CC effects on cells including induction or inhibition of collagen
 CC synthesis, fibrosis, differentiation responses, cell proliferative
 CC responses, and responses involving cell adhesion, migration, and
 CC extracellular matrices. These properties make the proteins potential
 CC agents for wound healing, reduction of fibrosis, and reduction of
 CC scar tissue formation. Chordin-related proteins may also be useful
 CC for advancement of the onset of fertility in sexually immature
 CC mammals, so as to increase the lifetime reproductive performance of
 CC domestic animals such as cows, sheep and pigs. Chordin-related
 CC proteins may also be useful in modulating hematopoiesis by inducing
 CC the differentiation of erythroid cells, for suppressing the
 CC development of gonadal tumors, or for augmenting the activity of
 CC BMPs. The proteins may also have value as a dietary supplement, or
 CC as a component of cell culture media.
 XX
 SQ Sequence 1036 AA;
 Query Match 60.9%; Score 78; DB 21; Length 1036;
 Best Local Similarity 70.6%; Pred. No. 0.041;
 Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 FAENETWVVDSTCTC 17
 Db 687 fvegetwnidstctc 703
 RESULT 11
 ID AAY53034 standard; Protein; 1036 AA.
 AC AAY53034;
 XX
 DT 29-FEB-2000 (first entry)
 XX
 DE Human secreted protein clone dj167_19 protein sequence SEQ ID NO:74.
 XX
 KW Human; secreted protein; nutritional; cytokine; cell proliferation;
 KW differentiation; immune stimulating; vaccine; suppression;
 KW hematopoiesis regulation; tissue growth; activin; inhibitor;
 KW chemotactic; chemokinetic; haemostatic; thrombolytic; receptor;
 KW ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
 KW tumour inhibition; gene therapy.
 XX

OS Homo sapiens.
XX WO9957132-A1.
PN
XX
PD 11-NOV-1999.
XX
XX 07-MAY-1999; 99WO-US09970.
XX PF
XX 07-MAY-1998; 98US-0084564.
PR 02-JUN-1998; 98US-0087645.
PR 22-JUL-1998; 98US-0093712.
PR 31-JUL-1998; 98US-0094935.
PR 10-AUG-1998; 98US-0095880.
PR 11-AUG-1998; 98US-0096068.
PR 06-MAY-1999; 99US-0096068.
XX
XX (GEMY) GENETICS INST INC.
PA
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Bowman MR;
PI DiBlasio-Smith E, Widom A;
XX
DR WPI: 2000-052937/04.
DR N-PSDB; AA233352.
XX
XX New polynucleotides encoding secreted human proteins, derived from
PT adult placenta, adult retina, fetal brain, fetal
XX
XX Claim 83; Page 426-429; 492pp; English.
XX
XX The present invention describes new human secreted proteins which were
CC isolated from adult placenta, adult retina, foetal brain, foetal kidney,
CC adult blood, adult brain, adult thyroid, adult bladder, adult neural
CC tissue, adult testes, and adult lymph node cDNA libraries. The human
CC secreted proteins, and the polynucleotides encoding them, are predicted
CC to have biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals. Suggested activities include nutritional activity, cytokine
CC and cell proliferation/differentiation activity, immune stimulating
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, and tumour inhibition
CC activity. The polynucleotides are also stated to be useful for gene
CC therapy. AA233316 to AA233373 encode human secreted proteins, and
CC AA232998 to AA233060 represent human secreted proteins, given in the
CC present invention.
XX
XX Sequence 1036 AA;
SQ

Query Match 50.9%; Score 78; DB 21; Length 1036;
Best Local Similarity 70.6%; Pred. No. 0.041;
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 FAENETWVDSCCTCTC 17
Db 587 fvegetwnidscctctc 703

RESULT 12
AAU07141
ID AAU07141 standard; Protein; 1036 AA.
XX AAU07141;
AC
XX 24-OCT-2001 (first entry)
DT
XX Human CRIM1 protein.
DE
XX CRIM-1; Human; human chromosome 2p21-16.3; ophthalmological;
KW neuroprotective; renal; osteopathic; dental; vulnerary; immunogen;
KW

KW antibody; gene therapy; neurodegenerative disease; eye disorder;
KW cataract; bone morphogenic protein; BMG; renal disease; bone abnormality;
KW tooth abnormality; wound; S52.
XX
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH Peptide 1..17
FT /label= Signal_peptide
FT Domain 1..901
FT /label= Ectodomain
FT /note= "This sequence is specifically claimed in claim
15"
FT Protein 18..1036
FT /label= Mature_CRIM1
FT Region 200..207
FT /note= "Conserved N-terminal motif"
FT Region 336..391
FT /label= CR_1
FT /note= "Cysteine rich repeat"
FT Region 403..456
FT /label= CR_2
FT /note= "Cysteine rich repeat"
FT Misc-difference 414
FT /note= "Encoded by GAC"
FT Region 608..662
FT /label= CR_3
FT /note= "Cysteine rich repeat"
FT Region 679..734
FT /label= CR_4
FT /note= "Cysteine rich repeat"
FT Region 753..808
FT /label= CR_5
FT /note= "Cysteine rich repeat"
FT Region 819..873
FT /label= CR_6
FT /note= "Cysteine rich repeat"
XX
XX WO200138519-A1.
XX 31-MAY-2001.
XX
XX 24-NOV-2000; 2000WO-AU01435.
XX
XX 26-NOV-1999; 99AU-0004348.
XX
XX (UYQU) UNIV QUEENSLAND.
XX
XX Little M, Yamada T, Holmes G, Georgas K, Kolle G, Wilkinson L;
PI
XX WPI: 2001-343951/36.
XX N-PSDB; AAS11601.
DR
DR
XX Nucleic acids from human chromosome 2p21-16.3 and the encoded peptide,
PT useful for preventing, diagnosing and treating e.g. eye disease,
PT especially cataract formation -
PT
XX Claim 11; Fig 1; 169pp; English.
XX
XX The invention relates to nucleic acids from human chromosome 2p21-16.3
CC and the encoded peptide (and mouse and chicken orthologues) that
CC comprises a PGECPLP group, an insulin-like growth factor binding protein
CC (IGFBP)-like domain, cysteine-rich domains, an RGD (undefined) group
CC and a transmembrane domain. The protein, e.g. CRIM1, interacts with
CC peptides of the transforming growth factor superfamily. A composition
CC comprising an expression construct comprising the nucleic acids of the
CC invention or a mimetic which antagonises or mimics an activity of a CRIM1
CC polypeptide may be used in a method for modulating the biological
CC activity of a polypeptide of the bone morphogenic protein (BMP) family.
CC In this way they may be used to prevent or treat an eye disease,
CC especially cataract formation. They may also be used to treat
CC neurodegenerative diseases, renal and kidney disease, bone and tooth
CC abnormalities, wounds and skin damage, e.g. by use of the nucleic acid in

CC gene therapy by using antibodies directed against CRIM1 polypeptides.
 CC The present sequence represents human CRIM1 (AKA S52).
 XX
 SQ Sequence 1036 AA;

Query Match 60.9%; Score 78; DB 22; Length 1036;
 Best Local Similarity 70.6%; Pred. No. 0.041;
 Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 FAENETWVVDCTCTC 17
 DB 687 fvegetwnidstctctc 703

RESULT 13

AAU12242
 ID AAU12242 standard; Protein; 1036 AA.

XX AC AAU12242;

XX DT 24-OCT-2001 (first entry)

XX DE Human PRO4330 polypeptide sequence.

XX KW Human secretory and transmembrane; PRO; mammalian; cancer; lung;
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
 KW adipocyte; A-peptide; factor VIIA; gene therapy.

XX OS Homo sapiens.

XX PN WO200140456-A2.

XX PD 07-JUN-2001.

XX PF 01-DEC-2000; 2000WO-US32678.

XX PR 01-DEC-1999; 99WO-US28301.

XX PR 01-DEC-1999; 99WO-US28634.

XX PR 02-DEC-1999; 99WO-US28551.

XX PR 02-DEC-1999; 99WO-US28564.

XX PR 02-DEC-1999; 99WO-US28565.

XX PR 09-DEC-1999; 99US-0170262.

XX PR 16-DEC-1999; 99WO-US30095.

XX PR 20-DEC-1999; 99WO-US30911.

XX PR 20-DEC-1999; 99WO-US30999.

XX PR 30-DEC-1999; 99WO-US31243.

XX PR 06-JAN-2000; 2000WO-US00277.

XX PR 11-FEB-2000; 2000WO-US00376.

XX PR 18-FEB-2000; 2000WO-US03565.

XX PR 18-FEB-2000; 2000WO-US04341.

XX PR 22-FEB-2000; 2000WO-US04342.

XX PR 24-FEB-2000; 2000WO-US04914.

XX PR 24-FEB-2000; 2000WO-US05004.

XX PR 01-MAR-2000; 2000WO-US05601.

XX PR 20-MAR-2000; 2000WO-US07377.

XX PR 21-MAR-2000; 2000WO-US07532.

XX PR 30-MAR-2000; 2000WO-US08439.

XX PR 17-MAY-2000; 2000WO-US13705.

XX PR 22-MAY-2000; 2000WO-US14042.

XX PR 30-MAY-2000; 2000WO-US14941.

XX PR 02-JUN-2000; 2000WO-US15264.

XX PR 10-NOV-2000; 2000WO-US30873.

XX PA (GETH) GENENTECH INC.

XX PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;

XX PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;

XX PI Smith V, Stewart TA, Tamas D, Watanabe CK, Wood WI, Zhang Z;

XX DW-WPI; 2001-408281/43.

DR N-PSDB; AAS21314.

XX Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical

XX Claim 12; Fig 142; 813pp; English.

XX AAU12172-AAU12446 represent novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.

XX SQ Sequence 1036 AA;

Query Match 60.9%; Score 78; DB 22; Length 1036;
 Best Local Similarity 70.6%; Pred. No. 0.041;
 Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 FAENETWVVDCTCTC 17
 DB 687 fvegetwnidstctctc 703

RESULT 14

AAU07142
 ID AAU07142 standard; Protein; 1037 AA.

XX AC AAU07142;

XX DT 24-OCT-2001 (first entry)

XX DE Mouse CRIM1 protein.

XX CRIM-1; Mouse; human chromosome 2p21-16.3; ophthalmological;
 KW neuroprotective; renal; osteopathic; dental; vulnary; immunogen;
 KW antibody; gene therapy; neurodegenerative disease; eye disorder;
 KW cataract; bone morphogenic protein; BMG; renal disease; bone abnormality;
 KW tooth abnormality; wound; S52.

XX OS Mus sp.

XX Key Location/Qualifiers
 FT Peptide 1..9
 FT Protein 10..1037
 FT Region 200..207
 FT Region 336..391
 FT Region /label= CR_1
 FT Region /note= "Cysteine rich repeat"
 FT Region 403..456
 FT Region /label= CR_2
 FT Region /note= "Cysteine rich repeat"

```

FT Region 608..662
FT /label= CR_3
FT /note= "Cysteine rich repeat"
FT Region 679..734
FT /label= CR_4
FT /note= "Cysteine rich repeat"
FT Region 753..808
FT /label= CR_5
FT /note= "Cysteine rich repeat"
FT Region 819..873
FT /label= CR_6
FT /note= "Cysteine rich repeat"
XX
XX WO200138519-A1.
XX
XX 31-MAY-2001.
XX
XX 24-NOV-2000; 2000WO-AU01435.
XX
XX 26-NOV-1999; 99AU-0004348.
XX (UYQU ) UNIV QUEENSLAND.
XX
XX Little M, Yamada T, Holmes G, Georgas K, Kolle G, Wilkinson L;
XX WPI; 2001-343951/36.
XX N-PSDB; AAS11602.
XX
XX Nucleic acids from human chromosome 2p21-16.3 and the encoded peptide,
XX useful for preventing, diagnosing and treating e.g. eye disease,
XX especially cataract formation -
XX
XX Claim 11; Fig 1; 169pp; English.
XX
XX The invention relates to nucleic acids from human chromosome 2p21-16.3
XX and the encoded peptide (and mouse and chicken orthologues) that
XX comprises a PGECPLP group, an insulin-like growth factor binding protein
XX (IGFBP)-like domain, cysteine-rich domains, an RGD (undefined) group
XX and a transmembrane domain. The protein, e.g. CRIM1, interacts with
XX peptides of the transforming growth factor superfamily. A composition
XX comprising an expression construct comprising the nucleic acids of the
XX invention or a mimetic which antagonises or mimics an activity of a CRIM1
XX polypeptide may be used in a method for modulating the biological
XX activity of a polypeptide of the bone morphogenic protein (BMP) family.
XX In this way they may be used to prevent or treat an eye disease,
XX especially cataract formation. They may also be used to treat
XX neurodegenerative diseases, renal and kidney disease, bone and tooth
XX abnormalities, wounds and skin damage, e.g. by use of the nucleic acid in
XX gene therapy by using antibodies directed against CRIM1 polypeptides.
XX The present sequence represents mouse CRIM1 (AKA S52).
XX
XX Sequence 1037 AA;
XX
XX Query Match 60.9%; Score 78; DB 22; Length 1037;
XX Best Local Similarity 70.6%; Pred. No. 0.041;
XX Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
XX
XX QY 1 FAENETWVDSCTTCTC 17
XX | | | | | | | | | |
XX Db 687 fvegetwnidsctqtc 703
XX
XX RESULT 15
XX AAU07143
XX ID AAU07143 standard; Protein; 1048 AA.
XX
XX AC AAU07143;
XX
XX XX 24-OCT-2001 (first entry)
XX DT
XX DE Chicken CRIM1 protein.
XX

```

```

KW CRIM-1; Chicken; human chromosome 2p21-16.3; ophthalmological;
KW neuroprotective; renal; osteopathic; dental; vulnary; immunogen;
KW antibody; gene therapy; neurodegenerative disease; eye disorder;
KW cataract; bone morphogenic protein; BMG; renal disease; bone abnormality;
XX tooth abnormality; wound; S52.
XX
XX Gallus gallus.
XX
XX Key Location/Qualifiers
XX Region 212..219
XX /note= "Conserved N-terminal motif"
XX Region 348..402
XX /label= CR_1
XX /note= "Cysteine rich repeat"
XX Region 415..468
XX /label= CR_2
XX /note= "Cysteine rich repeat"
XX Region 620..674
XX /label= CR_3
XX /note= "Cysteine rich repeat"
XX Region 691..746
XX /label= CR_4
XX /note= "Cysteine rich repeat"
XX Region 765..820
XX /label= CR_5
XX /note= "Cysteine rich repeat"
XX Region 831..885
XX /label= CR_6
XX /note= "Cysteine rich repeat"
XX
XX WO200138519-A1.
XX
XX 31-MAY-2001.
XX
XX 24-NOV-2000; 2000WO-AU01435.
XX
XX 26-NOV-1999; 99AU-0004348.
XX (UYQU ) UNIV QUEENSLAND.
XX
XX Little M, Yamada T, Holmes G, Georgas K, Kolle G, Wilkinson L;
XX WPI; 2001-343951/36.
XX N-PSDB; AAS11603.
XX
XX Nucleic acids from human chromosome 2p21-16.3 and the encoded peptide,
XX useful for preventing, diagnosing and treating e.g. eye disease,
XX especially cataract formation -
XX
XX Claim 11; Fig 1; 169pp; English.
XX
XX The invention relates to nucleic acids from human chromosome 2p21-16.3
XX and the encoded peptide (and mouse and chicken orthologues) that
XX comprises a PGECPLP group, an insulin-like growth factor binding protein
XX (IGFBP)-like domain, cysteine-rich domains, an RGD (undefined) group
XX and a transmembrane domain. The protein, e.g. CRIM1, interacts with
XX peptides of the transforming growth factor superfamily. A composition
XX comprising an expression construct comprising the nucleic acids of the
XX invention or a mimetic which antagonises or mimics an activity of a CRIM1
XX polypeptide may be used in a method for modulating the biological
XX activity of a polypeptide of the bone morphogenic protein (BMP) family.
XX In this way they may be used to prevent or treat an eye disease,
XX especially cataract formation. They may also be used to treat
XX neurodegenerative diseases, renal and kidney disease, bone and tooth
XX abnormalities, wounds and skin damage, e.g. by use of the nucleic acid in
XX gene therapy by using antibodies directed against CRIM1 polypeptides.
XX The present sequence represents chicken CRIM1 (AKA S52).
XX
XX Sequence 1048 AA;
XX
XX Query Match 60.9%; Score 78; DB 22; Length 1048;
XX Best Local Similarity 70.6%; Pred. No. 0.042;

```

Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 FAENETWVDSCTTCTC 17
| | | | | : | | | | |
Db 699 fvegetwnidstqctc 715

Search completed: January 31, 2002, 09:23:06
Job time: 196 sec

DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR001007; VWFC.
 DR InterPro; IPR003367; tsp_3.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF00090; tsp_1; 3.
 DR Pfam; PF02412; tsp_3; 9.
 DR Pfam; PF00093; vwc; 1.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00209; TSP1; 3.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50092; TSP1; 3.
 DR PROSITE; PS01208; VWFC; 1.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1172
 FT DOMAIN 19 232
 FT DOMAIN 318 375
 FT DOMAIN 381 432
 FT DOMAIN 437 493
 FT DOMAIN 494 548
 FT DOMAIN 549 589
 FT DOMAIN 590 647
 FT DOMAIN 648 692
 FT DOMAIN 725 760
 FT DOMAIN 761 783
 FT DOMAIN 784 819
 FT DOMAIN 820 842
 FT DOMAIN 843 880
 FT DOMAIN 881 916
 FT DOMAIN 917 952
 FT DOMAIN 953 1172
 FT SITE 928 930
 FT DISULFID 266 266
 FT DISULFID 270 270
 FT DISULFID 553 564
 FT DISULFID 558 574
 FT DISULFID 577 588
 FT DISULFID 594 610
 FT DISULFID 601 619
 FT DISULFID 622 646
 FT DISULFID 652 665
 FT DISULFID 659 678
 FT DISULFID 680 691
 FT CARBOHYD 151 151
 FT CARBOHYD 316 316
 FT CARBOHYD 330 330
 FT CARBOHYD 457 457
 FT CARBOHYD 584 584
 FT CARBOHYD 710 710
 FT CARBOHYD 1069 1069
 SQ SEQUENCE 1172 AA; 129955 MW; 2AC7BB230B44C6F5 CRC64;
 Query Match 100.08; Score 128; DB 1; Length 1172;
 Best Local Similarity 100.08; Pred. No. 5e-11;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FAENETWVVDSCCTCTCKKFKT 22
 Db 327 FAENETWVVDSCCTCTCKKFKT 348
 RESULT 2
 TSP2_MOUSE STANDARD; PRT; 1172 AA.
 AC Q03350;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE THROMBOSPONDIN 2 PRECURSOR.
 GN THBS2 OR TSP2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92147683; PubMed=1371115;
 RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
 RA Dixit V.M.;
 RT "Characterization of mouse thrombospondin 2 sequence and expression
 during cell growth and development.";
 RL J. Biol. Chem. 267:3274-3281(1992).
 RN [2]
 RP SEQUENCE OF 1-873 FROM N.A.
 RX MEDLINE=91302287; PubMed=1712771;
 RA Bornstein P., O'Rourke K., Wikstrom K., Wolf F.W., Katz R., Li P.,
 RA Dixit V.M.;
 RT "A second, expressed thrombospondin gene (Thbs2) exists in the mouse
 genome.";
 RL J. Biol. Chem. 266:12821-12824(1991).
 CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 LAMININ AND TYPE V COLLAGEN.
 CC -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
 CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
 CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
 CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 or send an email to license@isb-sib.ch).
 EMBL; L07803; AAA53064.1; -.
 EMBL; M64866; AAA40432.1; -.
 PIR; A42587; A42587.
 PIR; A39851; A39851.
 HSSP; P00740; 1IXA.
 MGD; MGI:98738; Thbs2.
 InterPro; IPR000561; EGF-like.
 InterPro; IPR001881; EGF_Ca.
 InterPro; IPR001791; Laminin_G.
 InterPro; IPR000884; TSP1.
 InterPro; IPR003129; TSPN.
 InterPro; IPR001007; VWFC.
 InterPro; IPR003367; tsp_3.
 Pfam; PF00008; EGF; 2.
 Pfam; PF02210; TSPN; 1.
 Pfam; PF00090; tsp_1; 3.
 Pfam; PF02412; tsp_3; 9.
 Pfam; PF00093; vwc; 1.
 SMART; SM00181; EGF; 3.
 SMART; SM00209; TSP1; 3.
 SMART; SM00210; TSPN; 1.
 SMART; SM00214; VWC; 1.
 PROSITE; PS00022; EGF_1; FALSE_NEG.
 PROSITE; PS01186; EGF_2; 1.
 PROSITE; PS50092; TSP1; 3.
 PROSITE; PS01208; VWFC; 1.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1172
 FT DOMAIN 19 232
 FT DOMAIN 318 375
 FT DOMAIN 381 432
 FT DOMAIN 437 493
 FT DOMAIN 494 548
 FT DOMAIN 549 589
 FT DOMAIN 590 647
 FT DOMAIN 648 692
 FT DOMAIN 725 760
 FT DOMAIN 761 783
 FT DOMAIN 784 819
 FT DOMAIN 820 842
 FT DOMAIN 843 880
 FT DOMAIN 881 916
 FT DOMAIN 917 952
 FT DOMAIN 953 1172
 FT SITE 928 930
 FT DISULFID 266 266
 FT DISULFID 270 270
 FT DISULFID 553 564
 FT DISULFID 558 574
 FT DISULFID 577 588
 FT DISULFID 594 610
 FT DISULFID 601 619
 FT DISULFID 622 646
 FT DISULFID 652 665
 FT DISULFID 659 678
 FT DISULFID 680 691
 FT CARBOHYD 151 151
 FT CARBOHYD 316 316
 FT CARBOHYD 330 330
 FT CARBOHYD 457 457
 FT CARBOHYD 584 584
 FT CARBOHYD 710 710
 FT CARBOHYD 1069 1069
 SQ SEQUENCE 1172 AA; 129955 MW; 2AC7BB230B44C6F5 CRC64;
 Query Match 100.08; Score 128; DB 1; Length 1172;
 Best Local Similarity 100.08; Pred. No. 5e-11;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FAENETWVVDSCCTCTCKKFKT 22
 Db 327 FAENETWVVDSCCTCTCKKFKT 348
 RESULT 2
 TSP2_MOUSE STANDARD; PRT; 1172 AA.
 AC Q03350;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)


```
FT DOMAIN 381 432 TSP TYPE-1 1.
FT DOMAIN 437 493 TSP TYPE-1 2.
FT DOMAIN 494 548 TSP TYPE-1 3.
FT DOMAIN 549 589 EGF-LIKE 1.
FT DOMAIN 590 647 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 648 692 EGF-LIKE 3.
FT DOMAIN 725 760 TSP TYPE-3 1.
FT DOMAIN 761 783 TSP TYPE-3 2.
FT DOMAIN 784 819 TSP TYPE-3 3.
FT DOMAIN 820 842 TSP TYPE-3 4.
FT DOMAIN 843 880 TSP TYPE-3 5.
FT DOMAIN 881 916 TSP TYPE-3 6.
FT DOMAIN 917 952 TSP TYPE-3 7.
FT DOMAIN 953 1172 C-TERMINAL.
FT SITE 928 930 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 266 266 INTERCHAIN (PROBABLE).
FT DISULFID 270 270 INTERCHAIN (PROBABLE).
FT DISULFID 553 564 BY SIMILARITY.
FT DISULFID 558 574 BY SIMILARITY.
FT DISULFID 577 588 BY SIMILARITY.
FT DISULFID 594 610 BY SIMILARITY.
FT DISULFID 601 619 BY SIMILARITY.
FT DISULFID 622 646 BY SIMILARITY.
FT DISULFID 652 665 BY SIMILARITY.
FT DISULFID 659 678 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1069 1069 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1172 AA; 129911 MW; 7CE8E4E8599822AB CRC64;

Query Match 100.0%; Score 128; DB 1; Length 1172;
Best Local Similarity 100.0%; Pred. No. 5e-11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FAENETWVDSCITCTCKKFKT 22
DB 327 FAENETWVDSCITCTCKKFKT 348

RESULT 3
TSP2_BOVIN STANDARD; PRT; 1170 AA.
ID TSP2_BOVIN Q28180;
AC Q95116; Q28180;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 2 PRECURSOR (CORTICOTROPIN-INDUCED SECRETED PROTEIN)
DE (CISP).
DE THB2 OR TSP2 OR TSP-2.
GN Bos taurus (Bovine).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Danik M., Chinn N.A., Lafaillade M., Keramidas M., Aguesse-Germon S.,
RA Penhoat A., Chen H., Mosher D., Chambaz E.M., Feige J.J.;
RA Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-522 FROM N.A.
RX MEDLINE=96331130; PubMed=8698834;
RA Lafaillade B., Pellerin S., Keramidas M., Danik M., Chambaz E.M.,
RA Feige J.J.;
RA *Opposite regulation of thrombospondin-1 and corticotropin-induced
RT secreted protein/thrombospondin-2 expression by adrenocorticotrophic
RT hormone in adrenocortical cells.*;
```

```
RL J. Cell. Physiol. 167:164-172(1996).
RN [3]
RP SEQUENCE OF 318-831 FROM N.A.
RC TISSUE-Aortic endothelium;
RA Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
RT "Cloning and sequencing of bovine thrombospondin stimulatory effect of
RT TGF-beta."
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X96540; CAA65385.1; -
CC EMBL; X87620; CAA60952.1; -
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001881; EGF-Ca.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR003129; TSPN.
CC InterPro; IPR001007; VWFC.
CC InterPro; IPR003367; tsp_3.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF02210; TSPN; 1.
CC Pfam; PF00090; tsp_1; 3.
CC Pfam; PF02412; tsp_3; 9.
CC Pfam; PF00093; vwc_1.
CC SMART; SM00181; EGF; 3.
CC SMART; SM00209; TSP1; 3.
CC SMART; SM00210; TSPN; 1.
CC SMART; SM00214; VWC; 1.
CC PROSITE; PS00022; EGF_1; FALSE_NEG.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS00092; TSP1; 3.
CC PROSITE; PS01208; VWFC; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1170
FT DOMAIN 19 232 THROMBOSPONDIN 2.
FT DOMAIN 318 375 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 379 430 VWFC.
FT DOMAIN 435 491 TSP TYPE-1 1.
FT DOMAIN 492 546 TSP TYPE-1 2.
FT DOMAIN 547 587 TSP TYPE-1 3.
FT DOMAIN 588 645 EGF-LIKE 1.
FT DOMAIN 646 690 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 723 758 EGF-LIKE 3.
FT DOMAIN 759 781 TSP TYPE-3 1.
FT DOMAIN 782 817 TSP TYPE-3 2.
FT DOMAIN 818 840 TSP TYPE-3 3.
FT DOMAIN 841 878 TSP TYPE-3 4.
FT DOMAIN 879 914 TSP TYPE-3 5.
FT DOMAIN 915 950 TSP TYPE-3 6.
FT DOMAIN 951 1170 TSP TYPE-3 7.
FT SITE CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 266 266 INTERCHAIN (PROBABLE).
FT DISULFID 270 270 INTERCHAIN (PROBABLE).
FT DISULFID 551 562 BY SIMILARITY.
FT DISULFID 556 572 BY SIMILARITY.
```

```

FT DISULFID 575 586 BY SIMILARITY.
FT DISULFID 592 608 BY SIMILARITY.
FT DISULFID 599 617 BY SIMILARITY.
FT DISULFID 620 644 BY SIMILARITY.
FT DISULFID 650 663 BY SIMILARITY.
FT DISULFID 657 676 BY SIMILARITY.
FT DISULFID 678 689 BY SIMILARITY.
FT DISULFID 151 151 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 455 455 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 582 582 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 708 708 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 936 936 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 535 535 A -> V (IN REF. 3).
FT CONFLICT 748 748 S -> T (IN REF. 3).
SQ SEQUENCE 1170 AA; 129862 MW; 9CF1BF55B89A051 CRC64;

Query Match 95.3%; Score 122; DB 1; Length 1170;
Best Local Similarity 95.5%; Pred. No. 3.7e-10;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FAENETWVVDSCCTCKKFKT 22
DB 327 FAENETWVVDSCCTCKKFKT 348

RESULT 4
TSP2_CHICK
ID TSP2_CHICK STANDARD; PRT; 1178 AA.
AC P35440;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 2 PRECURSOR.
GN THB2 OR TSP2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauromorpha; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxId=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91217026; PubMed=2022631;
RA Lawler J., Duquette M., Ferro P.;
RT "Cloning and sequencing of chicken thrombospondin.";
RL J. Biol. Chem. 266:8039-8043(1991).
CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 WFEC DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60853; AAA51437.1; -
DR PIR; A39804; A39804.
DR HSSP; P00740; 1IXA.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.

```

```

DR InterPro; IPR000884; TSP1.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; WFEC.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 8.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00001; EGF_Like; 1.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50092; TSP1; 3.
DR PROSITE; PS01208; WFEC; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT CHAIN ? 1178 THROMBOSPONDIN 2.
FT DOMAIN ? 322 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 324 381 WFEC.
FT DOMAIN 387 438 TSP TYPE-1 1.
FT DOMAIN 441 499 TSP TYPE-1 2.
FT DOMAIN 500 553 TSP TYPE-1 3.
FT DOMAIN 555 595 EGF-LIKE 1.
FT DOMAIN 596 653 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 654 698 EGF-LIKE 3.
FT DOMAIN 731 766 TSP TYPE-3 1.
FT DOMAIN 767 789 TSP TYPE-3 2.
FT DOMAIN 790 825 TSP TYPE-3 3.
FT DOMAIN 826 848 TSP TYPE-3 4.
FT DOMAIN 849 886 TSP TYPE-3 5.
FT DOMAIN 887 922 TSP TYPE-3 6.
FT DOMAIN 923 958 TSP TYPE-3 7.
FT DOMAIN 959 1178 C-TERMINAL.
FT SITE 934 935 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 559 570 BY SIMILARITY.
FT DISULFID 564 580 BY SIMILARITY.
FT DISULFID 583 594 BY SIMILARITY.
FT DISULFID 600 616 BY SIMILARITY.
FT DISULFID 607 625 BY SIMILARITY.
FT DISULFID 628 652 BY SIMILARITY.
FT DISULFID 658 671 BY SIMILARITY.
FT DISULFID 665 684 BY SIMILARITY.
FT DISULFID 686 697 BY SIMILARITY.
FT CARBOHYD 157 157 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 463 463 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 590 590 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 716 716 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1075 1075 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1178 AA; 131816 MW; F37E02F42C8717A2 CRC64;

Query Match 67.2%; Score 86; DB 1; Length 1178;
Best Local Similarity 61.9%; Pred. No. 5.8e-05;
Matches 13; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 FAENETWVVDSCCTCKKFKK 21
DB 333 FADSESIVDSCCTCKCQDSK 353

RESULT 5
ID TSP1_HUMAN STANDARD; PRT; 1170 AA.
AC P07996;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)

```

20-AUG-2001 (Rel. 40, Last annotation update)
THROMBOSPONDIN 1 PRECURSOR.
THBS1 OR TSP1 OR TSP.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE-Endothelial cells;
MEDLINE=87057617; PubMed=2430973;
Lawler J., Hynes R.O.;
"The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins.";
J. Cell Biol. 103:1635-1648(1986).
[2]
SEQUENCE FROM N.A.
MEDLINE=89139590; PubMed=2918029;
Hennessy S.W., Frazier B.A., Kim D.D., Deckwerth T.L., Baumgartel D.M., Rotwein P., Frazier W.A.;
"Complete thrombospondin mRNA sequence includes potential regulatory sites in the 3' untranslated region.";
J. Cell Biol. 108:729-736(1989).
[3]
SEQUENCE OF 1-397 FROM N.A.
MEDLINE=87157592; PubMed=3030396;
Kobayashi S., Eden-McCutchan F., Framson P., Bornstein P.;
"Partial amino acid sequence of human thrombospondin as determined by analysis of cDNA clones: homology to malarial circumsporozoite proteins.";
Biochemistry 25:8418-8425(1986).
[4]
SEQUENCE OF 1-374 FROM N.A.
MEDLINE=86287276; PubMed=3461443;
Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
"Characterization of a cDNA encoding the heparin and collagen binding domains of human thrombospondin.";
Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).
[5]
SEQUENCE OF 1-166 FROM N.A.
MEDLINE=89291870; PubMed=2544587;
Laherty C.D., Gierman T.M., Dixit V.M.;
"Characterization of the promoter region of the human thrombospondin gene. DNA sequences within the first intron increase transcription.";
J. Biol. Chem. 264:11222-11227(1989).
[6]
SEQUENCE OF 1028-1170 FROM N.A.
la Fleur M., Jobin C., Gauthier J., Kreis C.G.;
Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN, LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-V/BETA-3 AND ALPHA-IIb/BETA-3.
-!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
-!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
-!- SIMILARITY: CONTAINS 1 WFEC DOMAIN.
-!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
-!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
-!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M25631; AAA36741.1; -;
EMBL; X04665; CAA28370.1; -;
EMBL; X14787; CAA32889.1; -;
EMBL; J04835; AAA61178.1; -;

DR EMBL; M99425; AAB59366.1; -;
DR PIR; A05172; A05172.
DR PIR; A25812; A25812.
DR PIR; A26155; A26155.
DR PIR; A30140; A30140.
DR PIR; A34274; A34274.
DR HSSP; P35555; IEMO.
DR MIM; 188060; -;
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; WFEC.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 8.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00209; TSPI; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00092; TSPI; 3.
DR PROSITE; PS01208; WFEC; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1170 THROMBOSPONDIN 1.
FT DOMAIN 19 232 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 316 373 WFEC.
FT DOMAIN 379 430 TSP TYPE-1 1.
FT DOMAIN 432 491 TSP TYPE-1 2.
FT DOMAIN 495 548 TSP TYPE-1 3.
FT DOMAIN 549 587 EGF-LIKE 1.
FT DOMAIN 588 645 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 646 690 EGF-LIKE 3.
FT DOMAIN 723 758 TSP TYPE-3 1.
FT DOMAIN 759 781 TSP TYPE-3 2.
FT DOMAIN 782 817 TSP TYPE-3 3.
FT DOMAIN 818 840 TSP TYPE-3 4.
FT DOMAIN 841 878 TSP TYPE-3 5.
FT DOMAIN 879 914 TSP TYPE-3 6.
FT DOMAIN 915 950 TSP TYPE-3 7.
FT DOMAIN 951 1170 C-TERMINAL.
FT SITE 926 928 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 270 270 INTERCHAIN (PROBABLE).
FT DISULFID 274 274 INTERCHAIN (PROBABLE).
FT DISULFID 551 562 BY SIMILARITY.
FT DISULFID 556 572 BY SIMILARITY.
FT DISULFID 575 586 BY SIMILARITY.
FT DISULFID 592 608 BY SIMILARITY.
FT DISULFID 599 617 BY SIMILARITY.
FT DISULFID 620 644 BY SIMILARITY.
FT DISULFID 650 663 BY SIMILARITY.
FT DISULFID 676 676 BY SIMILARITY.
FT DISULFID 678 689 BY SIMILARITY.
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 84 84 T -> A (IN REF. 2, 3 AND 4).
FT CONFLICT 523 523 T -> A (IN REF. 2).
SQ SEQUENCE 1170 AA; 129412 MW; 69B3DE5AE3A395E CRC64;

Query Match 53.1%; Score 68; DB 1; Length 1170;
Best Local Similarity 50.0%; Pred. No. 0.023;
Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

```
OY 1 FAENETWVVDSTCTCKKFKT 22
: | | | | | | | | | |
Db 325 YRNEETWVDSCTECHQNSVT 346

RESULT 6
TSPI_MOUSE STANDARD: PRT: 1170 AA.
AC P35441;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 1 PRECURSOR.
GN THBS1 OR TSPI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=92128941; PubMed=1774063;
RA Lawler J., Duquette M., Ferro P., Copeland N.G., Gilbert D.J.,
RA Jenkins N.A.;
RT "Characterization of the murine thrombospondin gene.";
RL Genomics 11:587-600(1991).
[2]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=92147683; PubMed=1371115;
RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
RA Dixit V.M.;
RT "Characterization of mouse thrombospondin 2 sequence and expression
during cell growth and development.";
RL J. Biol. Chem. 267:3274-3281(1992).
[3]
RN RP SEQUENCE OF 1-490 FROM N.A.
RX MEDLINE=90375546; PubMed=2398070;
RA Bornstein P., Alf D., Devarayalu S., Framson P., Li P.;
RT "Characterization of the mouse thrombospondin gene and evaluation of
the role of the first intron in human gene expression.";
RL J. Biol. Chem. 265:16691-16698(1990).
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-
V/BETA-3 AND ALPHA-IIB/BETA-3.
CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 WFEC DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; M62470; AAA50611.1; JOINED.
DR EMBL; M62450; AAA50611.1; JOINED.
DR EMBL; M62451; AAA50611.1; JOINED.
DR EMBL; M62452; AAA50611.1; JOINED.
DR EMBL; M62453; AAA50611.1; JOINED.
DR EMBL; M62454; AAA50611.1; JOINED.
DR EMBL; M62455; AAA50611.1; JOINED.
DR EMBL; M62456; AAA50611.1; JOINED.
DR EMBL; M62457; AAA50611.1; JOINED.
DR EMBL; M62458; AAA50611.1; JOINED.
DR EMBL; M62459; AAA50611.1; JOINED.
DR EMBL; M62460; AAA50611.1; JOINED.
DR EMBL; M62461; AAA50611.1; JOINED.
DR EMBL; M62462; AAA50611.1; JOINED.
DR EMBL; M62463; AAA50611.1; JOINED.
DR EMBL; M62464; AAA50611.1; JOINED.
DR EMBL; M62465; AAA50611.1; JOINED.
DR EMBL; M62466; AAA50611.1; JOINED.
DR EMBL; M62467; AAA50611.1; JOINED.
DR EMBL; M62468; AAA50611.1; JOINED.
DR EMBL; M62469; AAA50611.1; JOINED.
DR EMBL; M87276; AAA53063.1; -
DR EMBL; J05605; AAA40431.1; -
DR EMBL; J05606; AAA40431.1; JOINED.
DR PIR; A40558; A40558.
DR PIR; B42587; B42587.
DR PIR; A37905; A37905.
DR HSP; P35555; IEMO.
DR MGP; MGI:98737; Thbs1.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF-Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR003129; TSPI.
DR InterPro; IPR001007; WFEC.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 8.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00209; TSPI; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VMC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00092; TSPI; 3.
DR PROSITE; PS01208; WFEC; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT CHAIN 1 18
FT DOMAIN 19 1170
FT DOMAIN 19 232
FT DOMAIN 316 373
FT DOMAIN 379 430
FT DOMAIN 435 491
FT DOMAIN 492 548
FT DOMAIN 549 587
FT DOMAIN 588 645
FT DOMAIN 646 690
FT DOMAIN 723 758
FT DOMAIN 759 781
FT DOMAIN 782 817
FT DOMAIN 818 840
FT DOMAIN 841 878
FT DOMAIN 879 914
FT DOMAIN 915 950
FT DOMAIN 951 1170
FT SITE 926 928
FT DISULFID 270 270
FT DISULFID 274 274
FT DISULFID 551 562
FT DISULFID 556 572
FT DISULFID 575 586
FT DISULFID 592 608
FT DISULFID 599 617
FT DISULFID 620 644
FT DISULFID 650 663
FT DISULFID 657 676
FT DISULFID 678 689
FT CARBOHYD 248 248
FT CARBOHYD 360 360
FT CARBOHYD 708 708
FT CARBOHYD 1067 1067
FT CONFLICT 1025 1025
SQ SEQUENCE 1170 AA; 129646 MW; 0443E493615E7F06 CRC64;
```

```
Query Match          53.1%; Score 68; DB 1; Length 1170;
Best Local Similarity 50.0%; Pred. No. 0.023;
Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 1 FAENETWVDSCCTTCKKFKT 22
    : ||||| ||| :
Db 325 YKNEETWVDSCTECHQNSVT 346

RESULT 7
TSPL_XENLA
ID TSPL_XENLA STANDARD; PRT; 1173 AA.
AC P35448;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 1 PRECURSOR.
GN THBS1 OR TSPL.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Dry L.A., Ramos J., Duquette M., Desimone D.W., Lawler J.;
RL Submitted (XXX-1993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-
CC V/BETA-3 AND ALPHA-1B/BETA-3 (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 WFEC DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04278; ; NOT_ANNOTATED_CDS.
DR HSSP; P00740; IIXA.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000884; TSPL.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; WFEC.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 8.
DR Pfam; PF00093; vwc; 1.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00001; EGF-like; 1.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; vwc; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50092; TSP1; 3.
DR PROSITE; PS01208; WFEC; 1.
DR KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
```

```
FT SIGNAL 1 22
FT CHAIN 23 1173
FT DOMAIN 23 235
FT DOMAIN 319 376
FT DOMAIN 382 433
FT DOMAIN 438 494
FT DOMAIN 495 546
FT DOMAIN 550 590
FT DOMAIN 591 648
FT DOMAIN 649 693
FT DOMAIN 726 761
FT DOMAIN 762 784
FT DOMAIN 785 820
FT DOMAIN 821 843
FT DOMAIN 844 881
FT DOMAIN 882 917
FT DOMAIN 918 953
FT DOMAIN 954 1173
FT SITE 929 931
FT DISULFID 554 565
FT DISULFID 559 575
FT DISULFID 578 589
FT DISULFID 595 611
FT DISULFID 602 620
FT DISULFID 623 647
FT DISULFID 653 666
FT DISULFID 660 679
FT DISULFID 681 692
FT CARBOHYD 155 155
FT CARBOHYD 158 158
FT CARBOHYD 250 250
FT CARBOHYD 363 363
FT CARBOHYD 705 705
FT CARBOHYD 711 711
FT CARBOHYD 1070 1070
SQ SEQUENCE 1173 AA; 130019 MW; A9F036D6516C0F24 CRC64;

Query Match          50.8%; Score 65; DB 1; Length 1173;
Best Local Similarity 55.6%; Pred. No. 0.063;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 5 ETWYVDSCCTTCTCKKFKT 22
    : ||||| ||| :
Db 332 DEWTVDSCTECTCQNSAT 349

RESULT 8
TSPL_BOVIN
ID TSPL_BOVIN STANDARD; PRT; 1170 AA.
AC Q28178; Q28179;
DT 01-NOV-1997 (Rel. 35, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 1 PRECURSOR.
GN THBS1 OR TSPL OR TSP-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HOLSTEIN; TISSUE-Tooth;
RX MEDLINE=98173773; PubMed=9507054;
RA Deno A., Yamashita K., Nagata T., Tsurumi C., Miwa Y., Kitamura S.,
RA Inoue H.;
RT "cDNA cloning of bovine thrombospondin 1 and its expression in
RT odontoblasts and predentin."
RL Biochim. Biophys. Acta 1382:17-22(1998).
RN [2]
RP SEQUENCE OF 1-18 AND 710-1170 FROM N.A.
RC TISSUE-Aortic endothelium;
```

RA Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC !- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 CC LAMININ, TYPE V COLLAGEN AND INTERGELIN ALPHA-V/BETA-1, ALPHA-
 CC V/BETA-3 AND ALPHA-IIB/BETA-3. MAY PLAY A ROLE IN DENTINOGENESIS
 CC AND/OR MAINTENANCE OF DENTIN AND DENTAL PULP.
 CC !- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
 CC !- TISSUE SPECIFICITY: ODONTOBLASTS.
 CC !- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC !- SIMILARITY: CONTAINS 1 WFEC DOMAIN.
 CC !- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC !- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
 CC !- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB005287; BAA21115.1; -;
 DR EMBL; X87618; CAA60950.1; -;
 DR EMBL; X87619; CAA60951.1; -;
 DR HSSP; P35555; IEMO.
 DR GlycoSuiteDB; Q28178; -;
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF-Ca.
 DR InterPro; IPR001791; Laminin-G.
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR001107; WFEC.
 DR InterPro; IPR003367; tsp_3.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF00090; tsp_1; 3.
 DR Pfam; PF02412; tsp_3; 16.
 DR Pfam; PF00093; vwc; 1.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00209; TSP1; 3.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; FALSE_NEG.
 DR PROSITE; PS01208; WFEC; 1.
 DR PROSITE; PS00092; TSP1; 3.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1170
 FT THROMBOSPONDIN 1.
 FT HEPARIN-BINDING (POTENTIAL).
 FT VWC.
 FT TSP TYPE-1 1.
 FT TSP TYPE-1 2.
 FT TSP TYPE-1 3.
 FT EGF-LIKE 1.
 FT EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
 FT EGF-LIKE 3.
 FT TSP TYPE-3 1.
 FT TSP TYPE-3 2.
 FT TSP TYPE-3 3.
 FT TSP TYPE-3 4.
 FT TSP TYPE-3 5.
 FT TSP TYPE-3 6.
 FT TSP TYPE-3 7.
 FT C-TERMINAL.
 FT CELL ATTACHMENT SITE (POTENTIAL).
 FT INTERCHAIN (PROBABLE).
 FT INTERCHAIN (PROBABLE).
 FT BY SIMILARITY.
 FT DISULFID 556

FT DISULFID 575 586 BY SIMILARITY.
 FT DISULFID 592 608 BY SIMILARITY.
 FT DISULFID 599 617 BY SIMILARITY.
 FT DISULFID 620 644 BY SIMILARITY.
 FT DISULFID 650 663 BY SIMILARITY.
 FT DISULFID 657 676 BY SIMILARITY.
 FT DISULFID 678 689 BY SIMILARITY.
 FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1085 1085 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 805 805 S -> G (IN REF. 2).
 SQ SEQUENCE 1170 AA; 129533 MW; 0DD6ADF3E5FA031A CRC64;

 Query Match 46.1%; Score 59; DB 1; Length 1170;
 Best Local Similarity 40.9%; Pred. No. 0.46; Mismatches 3; Indels 0; Gaps 0;
 Matches 9; Conservative 3;
 QY 1 FAENETWVVDSCCTCTCKRKFT 22
 DB 325 YTGDEWTVDSCTECRCQNSVT 346

 RESULT 9
 PSSP_HUMAN
 ID PSSP_HUMAN STANDARD; PRT; 114 AA.
 AC P08118; P11999;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROSTATE SECRETED SEMINAL PLASMA PROTEIN PRECURSOR (PSP-94) (BETA-
 DE MICROSEMINOPROTEIN) (SEMINAL PLASMA BETA-INHIBIN) (IMMUNOGLOBULIN
 DE BINDING FACTOR) (IGBF) (PROSTATE SECRETORY PROTEIN PSP94) (PN44).
 GN MSMB OR PRSP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87161231; PubMed=3829888;
 RA Mbikay M., Nolet S., Fournier S., Benjannet S., Chapdelaine P.,
 RA Paradis G., Dube J.Y., Tremblay R., Lazure C., Seidah N.G.,
 RA Chretien M.;
 RT "Molecular cloning and sequence of the cDNA for a 94-amino-acid
 RT seminal plasma protein secreted by the human prostate.";
 RL DNA 6:23-29(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91274357; PubMed=2054385;
 RA Nolet S., Mbikay M., Chretien M.;
 RT "Prostatic secretory protein psp94: gene organization and promoter
 RT sequence in rhesus monkey and human.";
 RL Biochim. Biophys. Acta 1089:247-249(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90211299; PubMed=2322265;
 RA Green C.B., Liu W.Y., Kwok S.C.M.;
 RT "Cloning and nucleotide sequence analysis of the human beta-
 RT microseminoprotein gene.";
 RL Biochem. Biophys. Res. Commun. 167:1184-1190(1990).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90073664; PubMed=2590204;
 RA Ulvsbaeck M., Lindstrom C., Weiber H., Abrahamsson P.-A., Lilja H.,
 RA Lundwall A.;
 RT "Molecular cloning of a small prostate protein, known as beta-
 RT microseminoprotein, PSP94 or beta-inhibin, and demonstration of
 RT transcripts in non-genital tissues.";
 RL Biochem. Biophys. Res. Commun. 164:1310-1315(1989).
 RN [5]

DR EMBL; X57934; CAAA1003.1; JOINED.
 DR EMBL; X57935; CAAA1003.1; JOINED.
 DR EMBL; M92161; AAA36903.1; -
 DR PIR; S16237; S16237.
 KW Sperm; Signal.
 FT SIGNAL 1 20 BY SIMILARITY.
 FT CHAIN 21 114 PROSTATE SECRETED SEMINAL PLASMA PROTEIN.
 SQ SEQUENCE 114 AA; 13079 MW; C07A4EB984B276C CRC64;
 Query Match 42.2%; Score 54; DB 1; Length 114;
 Best Local Similarity 50.0%; Pred. No. 0.29;
 Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
 QY 4 NETWVDSCTTCTCK 19
 DB 49 NSKWQTDNCERCYCK 64
 RESULT 12
 PSSP_PAPAN STANDARD; PRT; 114 AA.
 AC Q28767;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE PROSTATE SECRETED SEMINAL PLASMA PROTEIN PRECURSOR (PSP-94).
 GN MSMB OR PSP94.
 OS Papio anubis (olive baboon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Papio.
 OX NCBI_TaxID=9555;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Prostate;
 RX MEDLINE=97316893; PubMed=9174167;
 RA Xuan J.W., Wu D., Guo Y., Shum D.T., Mbikay M., Zhong R.,
 RA Chin J.L.;
 RT "Molecular cloning and gene expression analysis of PSP94 (prostate
 secretory protein of 94 amino acids) in primates.";
 RL DNA Cell Biol. 16:627-638(1997).
 CC -1- FUNCTION: INHIBITS THE SECRETION OF FSH BY PITUITARY CELLS (BY
 SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: SPERM SURFACE.
 CC -1- PTM: RICH IN DISULFIDE BONDS (POTENTIAL).
 CC -1- MISCELLANEOUS: SPECIFIC RECEPTORS FOR THIS PROTEIN ARE FOUND ON
 SPERMATOZOA AND IN THE PROSTATE (BY SIMILARITY).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U49786; AAB62726.1; -
 KW Sperm; Signal.
 FT SIGNAL 1 20 BY SIMILARITY.
 FT CHAIN 21 114 PROSTATE SECRETED SEMINAL PLASMA PROTEIN.
 SQ SEQUENCE 114 AA; 13013 MW; A08C837ED81F98ED CRC64;
 Query Match 40.6%; Score 52; DB 1; Length 114;
 Best Local Similarity 50.0%; Pred. No. 0.57;
 Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 QY 4 NETWVDSCTTCTC 17
 DB 49 NSKWQTDNCERACTC 62

RESULT 13
 MUC2_HUMAN STANDARD; PRT; 5179 AA.
 ID AC Q02817; Q14878;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).
 GN MUC2 OR SMUC.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Intestine;
 RX MEDLINE=94132002; PubMed=8300571;
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
 RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
 RT Identification of the amino terminus and overall sequence similarity
 to prepro-von Willebrand factor";
 RL J. Biol. Chem. 269:2440-2446(1994).
 RN [2]
 RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
 RC TISSUE=Colon;
 RX MEDLINE=93016075; PubMed=1400449;
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
 RA Kim Y.S.;
 RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
 both upstream and downstream of its central repetitive region.";
 RL J. Biol. Chem. 267:21375-21383(1992).
 RN [3]
 RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
 RX MEDLINE=91358717; PubMed=1885763;
 RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
 RA Petersen G.M., Kim Y.S.;
 RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
 and polymorphism.";
 RL J. Clin. Invest. 88:1005-1013(1991).
 CC -1- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
 OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
 PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
 AGENTS AT MUCOSAL SURFACES.
 CC -1- SUBUNIT: MULTIMERIC.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
 BRONCHUS, CERVIX AND GALL BLADDER.
 CC -1- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
 INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
 CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
 VARIES AMONG DIFFERENT ALLELES.
 CC -1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
 OF SILKWORM HEMOCYTIN.
 CC -1- SIMILARITY: CONTAINS 2 VWFC DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CFCK).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; L21998; AAB95295.1; -
 DR EMBL; M74027; AAA59875.1; -
 DR EMBL; M94131; AAA59163.1; -
 DR EMBL; M94132; AAA59164.1; -
 DR MIM; 158370; -
 DR InterPro; IPR000359; Cys_knot.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002400; GF_cysknot.
 DR InterPro; IPR001007; VWFC.

DR InterPro; IPR001846; Vwd.
DR Pfam; PF00007; Cys_knot; 1.
DR Pfam; PF00094; vwd; 4.
DR PRINTS; PR00438; GFCYSKNOT.
DR SMART; SM00214; VWC; 2.
DR SMART; SM00011; VWC_def; 2.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS01208; VWFC; 2.
KW Glycoprotein; Repeat; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 5179 MUCIN 2.
FT DOMAIN 1401 1747 APPROXIMATE REPEATS.
FT REPEAT 1401 1416 1.
FT REPEAT 1417 1432 2.
FT REPEAT 1433 1448 3.
FT REPEAT 1449 1464 4.
FT REPEAT 1465 1471 5.
FT REPEAT 1472 1478 6.
FT REPEAT 1479 1494 7A.
FT REPEAT 1495 1517 7B.
FT REPEAT 1518 1533 8A.
FT REPEAT 1534 1556 8B.
FT REPEAT 1557 1572 9A.
FT REPEAT 1573 1596 9B.
FT REPEAT 1597 1612 10A.
FT REPEAT 1613 1635 10B.
FT REPEAT 1636 1651 11A.
FT REPEAT 1652 1675 11B.
FT REPEAT 1676 1683 12.
FT REPEAT 1684 1699 13.
FT REPEAT 1700 1715 14.
FT REPEAT 1716 1731 15.
FT REPEAT 1732 1747 16.
FT DOMAIN 4815 4886 VWFC 1.
FT DOMAIN 4924 4991 VWFC 2.
FT DOMAIN 5075 5160 CTCK.
FT DISULFID 5075 5122 BY SIMILARITY.
FT DISULFID 5089 5136 BY SIMILARITY.
FT DISULFID 5098 5152 BY SIMILARITY.
FT DISULFID 5102 5154 BY SIMILARITY.
FT DISULFID ? 5159 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 770 770 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 894 894 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1154 1154 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1215 1215 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1230 1230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1246 1246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1787 1787 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1820 1820 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4339 4339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4351 4351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4362 4362 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4373 4373 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4422 4422 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4438 4438 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4502 4502 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4616 4616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4627 4627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4752 4752 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4787 4787 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4881 4881 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4888 4888 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4955 4955 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4970 4970 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5019 5019 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5038 5038 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5069 5069 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1351 1351 H -> L (IN REF. 3).
FT CONFLICT 1412 1412 T -> S (IN REF. 3).
FT CONFLICT 1449 1449 L -> P (IN REF. 3).
FT CONFLICT 1504 1504 M -> T (IN REF. 3).
FT CONFLICT 4192 4192 G -> S (IN REF. 2).
SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;
Query Match 40.6%; Score 52; DB 1; Length 5179;
Best Local Similarity 56.2%; Pred.No. 18;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 3 ENETWVDSCTTCTCK 18
DB 4421 ENETWVDSCTTCTCK 4436
RESULT 14
NEL2_MOUSE
ID NEL2_MOUSE STANDARD; PRT; 816 AA.
AC O61220;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN KINASE C-BINDING PROTEIN NEL2 PRECURSOR (NEL-LIKE PROTEIN 2)
DE (MEL91 PROTEIN).
GN NEL2 OR MEL91.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid:10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Elkins D.A., Rossi J.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP TSP N-TERMINAL DOMAIN.
RX MEDLINE-98153258; PubMed-9480764;
RA Beckmann G., Hanke J., Bork P., Reich J.;
RT "Merging extracellular domains; fold prediction for laminin G-like
RT and amino-terminal thrombospondin-like modules based on homology to
RT pentraxins.";
RL J. Mol. Biol. 275:725-730(1998).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 TSP N-TERMINAL DOMAIN.
CC -!- SIMILARITY: CONTAINS 5 VWFC DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U59230; AAB02924.1; ALT_INIT.
DR HSP; P00740; 11XA.
DR MGB; MGI:1858510; Nell2.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00093; vwc; 2.
DR SMART; SM00179; EGF_Ca; 3.
DR SMART; SM00001; EGF_like; 3.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.

```

DR SMART; SM00214; VWC; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS01208; VWC; 2.
KW Glycoprotein; EGF-like domain; Repeat; Signal.
FT SIGNAL 1 24
FT CHAIN 25 816
FT DOMAIN 30 258
FT DOMAIN 272 331
FT DOMAIN 332 396
FT DOMAIN 397 439
FT DOMAIN 440 481
FT DOMAIN 482 522
FT DOMAIN 521 553
FT DOMAIN 555 601
FT DOMAIN 602 637
FT DOMAIN 638 693
FT DOMAIN 698 756
FT DOMAIN 758 813
FT DISULFID 401 413
FT DISULFID 407 422
FT DISULFID 424 438
FT DISULFID 444 457
FT DISULFID 451 466
FT DISULFID 468 480
FT DISULFID 486 499
FT DISULFID 493 508
FT DISULFID 510 521
FT DISULFID 525 535
FT DISULFID 529 541
FT DISULFID 543 552
FT DISULFID 559 572
FT DISULFID 566 581
FT DISULFID 583 600
FT DISULFID 606 619
FT DISULFID 613 628
FT DISULFID 630 636
FT CARBOHYD 53 53
FT CARBOHYD 225 225
FT CARBOHYD 293 293
FT CARBOHYD 298 298
FT CARBOHYD 517 517
FT CARBOHYD 615 615
FT CARBOHYD 635 635
SQ SEQUENCE 816 AA; 91163 MW; 5BDD0A946F87E74D CRC64;

Query Match 40.2%; Score 51.5; DB 1; Length 816;
Best Local Similarity 47.1%; Pred. No. 4;
Matches 8; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 1 FAENETWVDSCTTCTC 17
: | : | | | |
Db 281 YRESHW-TDCKNCTC 296

RESULT 15
ID_NELL1_RAT STANDARD; PRT; 810 AA.
AC Q62919;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN KINASE C-BINDING PROTEIN NELL1 PRECURSOR (NEL-LIKE PROTEIN 1).
GN NELL1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX MEDLINE=20017976; PubMed=10548494;
RA Kuroda S., Oyasu M., Kawakami M., Kanayama N., Tanizawa K., Saito N.,
RA Abe T., Matsuhashi S., Ting K.;
RT "Biochemical characterization and expression analysis of neural
RL thrombospondin-1-like proteins NELL1 and NELL2.";
RL Biochem. Biophys. Res. Commun. 265:79-86(1999).
CC -1- SUBUNIT: HOMOTRIMER. BINDS TO PKC BETA-1.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: CONTAINS 1 TSP N-TERMINAL DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 VWC DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U48246; RAC72252.1; -.
DR HSSP; P07204; ZADX.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR001993; Mitoch_carrier.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; VWC.
DR Pfam; PF00008; EGF; 8.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00093; VWC; 5.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00001; EGF_like; 4.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS01208; VWC; 2.
KW Glycoprotein; EGF-like domain; Repeat; Signal.
FT SIGNAL 1 16
FT CHAIN 17 810
FT DOMAIN 81 230
FT DOMAIN 273 331
FT DOMAIN 335 390
FT DOMAIN 391 433
FT DOMAIN 434 475
FT DOMAIN 476 516
FT DOMAIN 515 547
FT DOMAIN 549 595
FT DOMAIN 596 631
FT DOMAIN 632 687
FT DOMAIN 692 750
FT DOMAIN 752 807
FT DISULFID 395 407
FT DISULFID 401 416
FT DISULFID 418 432
FT DISULFID 438 451
FT DISULFID 445 460
FT DISULFID 462 474
FT DISULFID 480 493
FT DISULFID 487 502
FT DISULFID 504 515
FT DISULFID 519 529
FT DISULFID 523 535
FT DISULFID 537 546
FT DISULFID 553 566
FT DISULFID 560 575
FT DISULFID 577 594

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:36:15 ; Search time 90.11 Seconds
(without alignments)
35.712 Million cell updates/sec

Title: US-09-536-087-8

Perfect score: 128

Sequence: 1 FAENETWVDSCTCTCKKFKT 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	60.9	925	4 Q9H318	Q9H318 homo sapien
2	78	60.9	1028	11 Q9JLL0	Q9JLL0 mus musculu
3	78	60.9	1036	4 Q9NZV1	Q9NZV1 homo sapien
4	70	54.7	2327	13 Q9IBG7	Q9IBG7 xenopus lae
5	68	53.1	376	4 Q15667	Q15667 homo sapien
6	64	50.0	1823	5 Q26638	Q26638 strongyloce
7	63	49.2	1376	5 Q26637	Q26637 strongyloce
8	61.5	48.0	719	10 Q9C7V1	Q9C7V1 arabidopsis
9	61.5	48.0	3290	10 Q9SH73	Q9SH73 arabidopsis
10	60	46.9	909	5 Q17429	Q17429 caenorhabdi
11	57	44.5	1496	4 Q92626	Q92626 homo sapien
12	57	44.5	3198	5 Q26639	Q26639 strongyloce
13	56.5	44.1	301	5 Q9VWP6	Q9VWP6 drosophila
14	56.5	44.1	314	5 Q9BLX1	Q9BLX1 drosophila
15	55	43.0	114	6 Q97949	Q97949 saguinus oe
16	55	43.0	3843	5 Q9U5D0	Q9U5D0 drosophila
17	55	43.0	3843	5 Q9VU94	Q9VU94 drosophila
18	54.5	42.6	5374	11 Q99ND0	Q99ND0 mus musculu
19	54	42.2	462	11 Q9CXM8	Q9CXM8 mus musculu

20	53	41.4	644	4	O14985	O14985 homo sapien
21	53	41.4	845	4	Q9UE28	Q9UE28 homo sapien
22	53	41.4	879	4	O15494	O15494 homo sapien
23	53	41.4	1527	5	Q9VZZ4	Q9VZZ4 drosophila
24	53	41.4	1535	5	Q23991	Q23991 drosophila
25	52.5	41.0	55	5	Q9U7C1	Q9U7C1 entamoeba i
26	52.5	41.0	220	5	O96338	O96338 entamoeba i
27	52	40.6	449	4	O9HBC5	O9HBC5 homo sapien
28	51.5	40.2	196	4	O93043	O93043 homo sapien
29	51.5	40.2	390	4	O00447	O00447 homo sapien
30	51.5	40.2	1378	5	O97405	O97405 halotidis di
31	51	39.8	665	5	Q9W2H2	Q9W2H2 drosophila
32	51	39.8	751	5	O9GYX3	O9GYX3 drosophila
33	49.5	38.7	725	11	O9CV93	O9CV93 mus musculu
34	49	38.3	72	12	P90132	P90132 human immun
35	49	38.3	78	11	Q9QXJ3	Q9QXJ3 rattus norv
36	49	38.3	1329	5	Q9BMB0	Q9BMB0 caenorhabdi
37	48.5	37.9	336	5	Q9VPC4	Q9VPC4 drosophila
38	48	37.5	109	5	O9NC78	O9NC78 leishmania
39	48	37.5	244	5	O9NE80	O9NE80 leishmania
40	47	36.7	269	11	O64037	O64037 rattus norv
41	47	36.7	364	2	O33303	O33303 mycobacteri
42	47	36.7	364	5	Q9VDM7	Q9VDM7 drosophila
43	47	36.7	3026	5	Q26030	Q26030 plasmodium
44	47	36.7	4845	11	O88738	O88738 mus musculu
45	46.5	36.3	2108	13	Q98U19	Q98U19 gallus gall

ALIGNMENTS

RESULT 1
Q9H318 PRELIMINARY; PRT; 925 AA.
ID Q9H318
AC Q9H318;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CRIM1 PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominiade; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kolle G., Georgas K., Holmes G.P., Little M.H., Yamada T.;
RT "CRIM1, a novel gene encoding a cysteine-rich repeat protein, is developmentally regulated and implicated in vertebrate CNS development and organogenesis."
RL Mech. Dev. 0:0-0(2000).
DR EMBL; AF168681; AAC37011.1; .
DR InterPro; IPR001007; VWFC.
DR Pfam; PF00093; VWC; 6.
DR SMART; SM00214; VWC; 7.
DR PROSITE; PS01208; VWFC; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 925 AA; 102089 MW; D397E079637DCDFD CRC64;

Query Match 60.9%; Score 78; DB 4; Length 925;
Best Local Similarity 70.6%; Pred. NO. 0.00057;
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 FAENETWVDSCTTCTC. 17
| | | | | : | | | | |
Db 576 FVEGETWNIDSCTQCTC 592

RESULT 2
Q9JLL0 PRELIMINARY; PRT; 1028 AA.
ID Q9JLL0
AC Q9JLL0;
DT 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CYSTEINE-RICH REPEAT-CONTAINING PROTEIN CRIM1 PRECURSOR (FRAGMENT).
GN CRIM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20108580; PubMed=10642437;
RA Kollie G., Georgasi K., Holmes G.P., Little M.H., Yamada T.;
RT "CRIM1, a novel gene encoding a cysteine-rich repeat protein, is
RT developmentally regulated and implicated in vertebrate CNS development
RT and organogenesis.";
RL Mech. Dev. 90:181-193(2000).
DR EMBL: AF168680; AAF34410.1; -.
DR MGD: MGI:1354756; Crim1.
DR InterPro; IPR000867; IGFBP.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF00093; vwc; 6.
DR SMART; SM00121; IB; 1.
DR SMART; SM00214; VWC; 6.
DR PROSITE; PS01208; VWFC; 6.
KW Signal.
FT NON_TER 1 1
FT SIGNAL <1 25
SQ SEQUENCE 1028 AA; 113159 MW; 743058AA481D5ED8 CRC64;

Query Match 60.9%; Score 78; DB 11; Length 1028;
Best Local Similarity 70.6%; Pred. No. 0.00062;
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 FAENETWVVDSCCTCTC 17
Db 678 FVEGETWNIDSCCTCTC 694

RESULT 3
Q9NZV1 PRELIMINARY; PRT; 1036 AA.
ID Q9NZV1
AC Q9NZV1
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CYSTEINE-RICH REPEAT-CONTAINING PROTEIN S52 PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20108580; PubMed=10642437;
RA Kollie G., Georgasi K., Holmes G.P., Little M.H., Yamada T.;
RT "CRIM1, a novel gene encoding a cysteine-rich repeat protein, is
RT developmentally regulated and implicated in vertebrate CNS development
RT and organogenesis.";
RL Mech. Dev. 90:181-193(2000).
DR EMBL: AF167706; AAF34409.1; -.
DR InterPro; IPR000867; IGFBP.
DR InterPro; IPR000169; Thiolprot_act_site.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF00093; vwc; 6.
DR SMART; SM00121; IB; 1.
DR SMART; SM00214; VWC; 6.
DR PROSITE; PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
DR PROSITE; PS01208; VWFC; 6.
KW Signal.
FT SIGNAL 1 34
SQ SEQUENCE 1036 AA; 113737 MW; 10CBF02A5C579C27 CRC64;

Query Match 60.9%; Score 78; DB 4; Length 1036;
Best Local Similarity 70.6%; Pred. No. 0.00063;
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 FAENETWVVDSCCTCTC 17
Db 687 FVEGETWNIDSCCTCTC 703

RESULT 4
Q9IBG7 PRELIMINARY; PRT; 2327 AA.
ID Q9IBG7
AC Q9IBG7
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE KIELIN.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=FLOOR PLATE;
RX MEDLINE=20266358; PubMed=10779551;
RA Matsui M., Mizusaki K., Nakatani J., Nakanishi S., Sasai Y.;
RT "Xenopus kielin: A dorsalizing factor containing multiple chordin-type
RT repeats secreted from the embryonic midline.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:5291-5296(2000).
DR EMBL: AB026192; BAA95483.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002919; TIL.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001846; Vwd.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF01826; TIL; 1.
DR Pfam; PF00093; vwc; 22.
DR Pfam; PF00094; vwd; 1.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 27.
DR SMART; SM00011; VWC_def; 1.
DR SMART; SM00216; VWD; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR PROSITE; PS01208; VWFC; 20.
SQ SEQUENCE 2327 AA; 255800 MW; 0293109329209983 CRC64;

Query Match 54.7%; Score 70; DB 13; Length 2327;
Best Local Similarity 64.7%; Pred. No. 0.022;
Matches 11; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 FAENETWVVDSCCTCTC 17
Db 1937 FLSNEHWQVDECTACTC 1953

RESULT 5
Q15667 PRELIMINARY; PRT; 376 AA.
ID Q15667
AC Q15667
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE THROMBOSPONDIN (TSP) PRECURSOR (FRAGMENT).
GN THBS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86287276; PubMed=3461443;

```

RA Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
RT *Characterization of a cDNA encoding the heparin and collagen binding
RL domains of human thrombospondin.*;
RN Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92278426; PubMed=1350660;
RA Bursell T.R., Barnes T.M.;
RT "Introns in sequence tags.";
RL Nature 357:367-368(1992).
DR EMBL; M14326; AAA61237.1; -.
DR InterPro; IPR001007; VFNC.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF00093; vwc; 1.
DR Pfam; PF02210; TSPN; 1.
DR PROSITE; PS01208; VFNC; 1.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 >376 POTENTIAL.
FT NON_TER 376
SQ SEQUENCE 376 AA; 41329 MW; 313B62C0D253BA2B CRC64;

Query Match 53.1%; Score 68; DB 4; Length 376;
Best Local Similarity 50.0%; Pred. No. 0.009;
Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 FAENETWVDSCTTCTCKKFKT 22
Db : ||| ||||| | :
325 YRNEETWVDSCTCHCQNSVT 346

RESULT 6
ID Q26638 PRELIMINARY; PRT; 1823 AA.
AC Q26638;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 2 ALPHA FIBRILLAR COLLAGEN (FRAGMENT).
GN COL2ALPHA.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96096722; PubMed=8529669;
RA Exposito J.Y., Boute N., Deleage G., Garrone R.;
RT *Characterization of two genes coding for a similar four-cysteine
RT motif of the amino-terminal propeptide of a sea urchin fibrillar
RT collagen.*;
RL Eur. J. Biochem. 234:59-65(1995).
DR EMBL; X89806; CAA61929.1; -.
DR InterPro; IPR001007; VFNC.
DR Pfam; PF00093; vwc; 1.
DR PROSITE; PS01208; VFNC; UNKNOWN_1.
DR SMART; SM00214; VWC; 1.
FT NON_TER 1823
SQ SEQUENCE 1823 AA; 200206 MW; 0C11AABDF5F35D00 CRC64;

Query Match 50.0%; Score 64; DB 5; Length 1823;
Best Local Similarity 45.5%; Pred. No. 0.15;
Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 FAENETWVDSCTTCTCKKFKT 22
Db : ||| ||||| | :
56 YLHGESWSVDECTTCEDNATT 77

```

```

RESULT 7
ID Q26637 PRELIMINARY; PRT; 1376 AA.
AC Q26637;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 5 ALPHA FIBRILLAR COLLAGEN (FRAGMENT).
GN COL5ALPHA.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96096722; PubMed=8529669;
RA Exposito J.Y., Boute N., Deleage G., Garrone R.;
RT *Characterization of two genes coding for a similar four-cysteine
RT motif of the amino-terminal propeptide of a sea urchin fibrillar
RT collagen.*;
RL Eur. J. Biochem. 234:59-65(1995).
DR EMBL; X89800; CAA61928.1; -.
DR EMBL; X89801; CAA61928.1; JOINED.
DR EMBL; X89802; CAA61928.1; JOINED.
DR EMBL; X89803; CAA61928.1; JOINED.
DR EMBL; X89804; CAA61928.1; JOINED.
DR EMBL; X89805; CAA61928.1; JOINED.
DR InterPro; IPR001007; VFNC.
DR Pfam; PF00093; vwc; 1.
DR PROSITE; PS01208; VFNC; UNKNOWN_1.
DR SMART; SM00214; VWC; 1.
FT NON_TER 1
SQ SEQUENCE 1376 AA; 151182 MW; AF134036781FAAC6 CRC64;

Query Match 49.2%; Score 63; DB 5; Length 1376;
Best Local Similarity 45.5%; Pred. No. 0.17;
Matches 10; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 1 FAENETWVDSCTTCTCKKFKT 22
Db : ||| ||||| | :
23 YLHGEWKVDECTTCACDNATT 44

RESULT 8
ID Q9C7V1 PRELIMINARY; PRT; 719 AA.
AC Q9C7V1;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL 83.2 KDA PROTEIN.
GN F15H21.15.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,

```

```

RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820(2000).
DR EMBL; AC066689; AAG51711.1; -.
KW Hypothetical protein.
SQ SEQUENCE 719 AA; 83215 MW; D7C7B12283729857 CRC64;

Query Match 48.0%; Score 61.5; DB 10; Length 719;
Best Local Similarity 45.5%; Pred. No. 0.16;
Matches 10; Conservative 7; Mismatches 4; Indels 1; Gaps 1;

QY 2 AENETWVVD-SCTTCTCKKFKT 22
Db 629 SEKEWIVQLNVSTCTCRKFQ 650
:| | | | | : : |||||:| | | |
| | | | | : : |||||:| | | |

RESULT 9
Q9SH73 PRELIMINARY; PRT; 3290 AA.
AC Q9SH73;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE F22C12.1.
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Khan S., Brooks S., Buehler E., Chao Q., Dunn P., Kim C.,
RA Walker M., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,
RA Hansen N.F., Huizar L., Kremenetskaia I., Lenz C., Li J., Liu S.,
RA Luros S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G.,
RA Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F22C12 from chromosome
RT 1.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007764; AAF4584.1; -.
DR InterPro; IPR000901; CPSase.
DR PROSITE; PS00867; CPSASE_2; UNKNOWN_1.
SQ SEQUENCE 3290 AA; 377026 MW; D227EEB20993C5FB CRC64;

Query Match 48.0%; Score 61.5; DB 10; Length 3290;
Best Local Similarity 45.5%; Pred. No. 0.62;
Matches 10; Conservative 7; Mismatches 4; Indels 1; Gaps 1;

QY 2 AENETWVVD-SCTTCTCKKFKT 22
Db 3200 SEKEWIVQLNVSTCTCRKFQ 3221
:| | | | | : : |||||:| | | |
| | | | | : : |||||:| | | |

RESULT 10
Q17429 PRELIMINARY; PRT; 909 AA.
AC Q17429;
DT 01-JAN-1999 (TrEMBLrel. 09, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE B0024.14 PROTEIN.
CN B0024.14.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

```

```

OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; ZL1178; CAA94886.2; -.
DR EMBL; AL021478; CAA94886.2; JOINED.
DR HSSP; P15358; 1SKZ.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF00093; vwc; 4.
DR SMART; SM00214; VWC; 8.
DR PROSITE; PS01208; VWFC; UNKNOWN_1.
SQ SEQUENCE 909 AA; 99832 MW; 489CFCCFF4C39F1AA CRC64;

Query Match 46.9%; Score 60; DB 5; Length 909;
Best Local Similarity 44.4%; Pred. No. 0.34;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 FAENETWVVD-SCTTCTCK 18
Db 544 FTDGETWQLAPCVSCTCR 561
| | | | | : | : | | |
| | | | | : | : | | |

RESULT 11
Q92626 PRELIMINARY; PRT; 1496 AA.
ID Q92626
AC Q92626;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MYELOBLAST KIAA0230 (FRAGMENT).
GN KIAA0230.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-BONE MARROW;
RC MEDLINE=97191544; PubMed=90395052;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-MELANOMA;
RX MEDLINE=95048383; PubMed=7959781;
RA Weller S.R., Taylor S.M., Deans R.J., Kan-Mitchell J., Mitchell M.S.,
RA Trent J.M.;
RT "Assignment of a human melanoma associated gene MG50 (D2S448) to
RT chromosome 2p25.3 by fluorescence in situ hybridization.";
RL Genomics 22:243-244(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-MELANOMA;
RA Mitchell M.S., Kan-Mitchell J., Minev B., Edman C., Deans R.J.;
RT "Identification of a novel melanoma gene (MG50) - likely the gene for
RT IL-1 receptor antagonist - which encodes epitopes recognized by human
RT cytolytic T lymphocytes.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
CC EMBL; D86983; BAA13219.1; -.
DR EMBL; AF200348; AAF06354.1; -.
DR HSSP; P05164; LXP.
DR InterPro; IPR002007; Anim_peroxidase.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.

```


DR InterPro: IPR001611; LRR.
 DR InterPro: IPR000483; LRR_Cterm.
 DR InterPro: IPR000372; LRR_Nterm.
 DR InterPro: IPR003591; LRR_Typ.
 DR InterPro: IPR001536; Peroxidase_3.
 DR InterPro: IPR002016; Peroxidase.
 DR InterPro: IPR001007; WFC.
 DR Pfam: PF00047; Ig; 4.
 DR Pfam: PF00560; LRR; 5.
 DR Pfam: PF01463; LRRCT; 1.
 DR Pfam: PF00141; peroxidase; 1.
 DR Pfam: PF00093; vwc; 1.
 DR PRINTS: PR00457; ANPEROXIDASE.
 DR SMART: SM00408; IGC2; 4.
 DR SMART: SM00082; LRRCT; 1.
 DR SMART: SM00013; LRRNT; 1.
 DR SMART: SM00369; LRR_Typ; 4.
 DR SMART: SM00214; WVC; 1.
 DR PROSITE: PS01208; WFC; UNKNOWN_1.
 FT NON_TER 1
 SQ SEQUENCE 1496 AA; 167209 MW; E9B9A7069BF1ABFF CRC64;

Query Match 44.5%; Score 57; DB 4; Length 1496;
 Best Local Similarity 52.9%; Pred. No. 1.5;
 Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 AENETWVDSCTTCTCK 18
 | | | | | | | |
 DB 1441 ANNTKWKDACECK 1457

RESULT 12
 Q26639 PRELIMINARY; PRT; 3198 AA.
 AC Q26639
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ALPHA-2 COLLAGEN.
 GN COL2ALPHA.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
 OC Strongylocentrotus.
 OX NCBI_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92348411; PubMed=1639795;
 RA Exposito J.Y., D'Alessio M., Solursh M., Ramirez F.;
 RT "Sea urchin collagen evolutionarily homologous to vertebrate pro-alpha
 2(I) collagen.";
 RL J. Biol. Chem. 267:15559-15562(1992).
 RN [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE=92381062; PubMed=1380962;
 RA Exposito J.Y., D'Alessio M., Ramirez F.;
 RT "Novel amino-terminal propeptide configuration in a fibrillar
 procollagen undergoing alternative splicing.";
 RL J. Biol. Chem. 267:17404-17408(1992).
 DR EMBL: M92041; AAA30040.1; -.
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR000885; Fib.collagen_C.
 DR InterPro: IPR001007; WFC.
 DR Pfam: PF00093; vwc; 1.
 DR Pfam: PF01391; Collagen; 17.
 DR Pfam: PF01410; COLFI; 1.
 DR ProDom: PD002078; Fib.collagen_C; 1.
 DR PROSITE: PS01208; WFC; UNKNOWN_1.
 DR SMART: SM00038; COLFI; 1.
 DR SMART: SM00214; WVC; 1.
 SQ SEQUENCE 3198 AA; 331465 MW; 114359E69542DCB2 CRC64;

Query Match 44.5%; Score 57; DB 5; Length 3198;
 Best Local Similarity 61.5%; Pred. No. 3;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 ETWVDSCTTCTC 17
 | | | | | | | |
 DB 60 ESWVDECTCSC 72

RESULT 13
 Q9VWP6 PRELIMINARY; PRT; 301 AA.
 ID Q9VWP6
 AC Q9VWP6;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CG7103 PROTEIN.
 GN CG7103.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkov D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003510; AAF48892.2; -.
 DR FlyBase: FBgn0030964; CG7103.
 DR InterPro: IPR000072; PDGF.
 DR Pfam: PF00341; PDGF; 1.
 DR ProDom: PD001629; PDGF; 1.
 DR SMART: SM00141; PDGF; 1.
 DR PROSITE: PS0278; PDGF_2; 1.

SQ SEQUENCE 301 AA; 33013 MW; 4C45D988AAC9A5A9 CRC64;

Query Match 44.1%; Score 56.5; DB 5; Length 301;
Best Local Similarity 58.8%; Pred. No. 0.44; 3; Indels 1; Gaps 1;
Matches 10; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

Qy 2 AENETWVVDSCCTCTCK 18
|||: || ||: || ||:
Db 261 AENKYWVDNC-TCVCR 276

RESULT 14

Q9BLX1 ID Q9BLX1 PRELIMINARY; PRT; 314 AA.
AC Q9BLX1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PDGF/VEGF-LIKE PROTEIN PRECURSOR.
GN CG7103.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP. SEQUENCE FROM N.A.
RA Heino T.I., Wahliroem G., Kaerpaenen T., Pulkkinen M., Alitalo K.,
RA Roos C.;
RT "Drosophila PDGF/VEGF receptor homolog is specifically expressed in
RT hemocytes.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ401391; CAC24699.1; -
KW Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 314 PDGF/VEGF-LIKE PROTEIN.
SQ SEQUENCE 314 AA; 34482 MW; DE777E7B2BD5EDD CRC64;

Query Match 44.1%; Score 56.5; DB 5; Length 314;
Best Local Similarity 58.8%; Pred. No. 0.46; 3; Indels 1; Gaps 1;
Matches 10; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

Qy 2 AENETWVVDSCCTCTCK 18
|||: || ||: || ||:
Db 249 AENKYWVDNC-TCVCR 264

RESULT 15

O97949 ID O97949 PRELIMINARY; PRT; 114 AA.
AC O97949;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE BETA-MICROSEMINOPROTEIN PRECURSOR.
GN MSP-J1.
OS Saguinus oedipus (Cotton-top tamarin).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Saguinus.
OX NCBI_TaxID=9490;
RN [1]
RP. SEQUENCE FROM N.A.
RA Lundwall A., Makinen M.;
RT "Molecular cloning of genes encoding beta-microseminoprotein from the
RT cotton-top tamarin.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ010155; CAB38123.1; -
DR EMBL; AJ010156; CAB38123.1; JOINED.
DR EMBL; AJ010157; CAB38123.1; JOINED.
KW Signal.
FT SIGNAL 0 19 POTENTIAL.

SQ SEQUENCE 114 AA; 12746 MW; F3F05B013445BAD4 CRC64;

Query Match 43.0%; Score 55; DB 6; Length 114;
Best Local Similarity 43.8%; Pred. No. 0.32; 5; Indels 0; Gaps 0;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 4 AENETWVVDSCCTCTCK 19
|||: || ||: || ||:
Db 49 NSKWRTDNCDSCTCRE 64

Search completed: January 31, 2002, 09:36:16
Job time: 785 sec

THIS PAGE BLANK (uspro)

QY	1	FAENETWVD	SCTTCTCKK	FKT	22
Db	327	FAENETWVD	SCTTCTCKK	FKT	348

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/01652
; FILING DATE: 19930222
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/841,656
; FILING DATE: 24-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/464,369
; FILING DATE: 12-JAN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Fentress, Susan B.
; REGISTRATION NUMBER: 31,327
; REFERENCE/DOCKET NUMBER: 92005-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312)-456-8000
; TELEFAX: (312)-456-7776
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 239 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; PCT-US93-01652-1

Query Match 53.1%; Score 68; DB 5; Len
Best Local Similarity 50.0%; Pred. NO. 0.022;
Matches 11; Conservative 2; Mismatches 9;

QY 1 EAENETWVVDSCCTCTCKKFKT 22
Db 14 YRNEETWVDSCTCHCQNSVT 35

RESULT 4
US-08-985-526-3
; Sequence 3, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Connolly, Bove, Lodge, & Hutzel
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mc Morrow Jr., Robert G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613

```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 441 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-985-526-3

Query Match 53.1%; Score 68; DB 3; Length 441;
Best Local Similarity 50.0%; Pred. No. 0.04;
Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 FAENETWVDSCCTCKKFKT 22
; : ||| ||||| :
Db 248 YRNEETWVDSCTECHQNSVT 269

RESULT 5
US-08-313-288B-20
; Sequence 20, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and AviHu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1170 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-313-288B-20

Query Match 53.1%; Score 68; DB 1; Length 1170;
Best Local Similarity 50.0%; Pred. No. 0.1;
Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 FAENETWVDSCCTCKKFKT 22
; : ||| ||||| :
Db 325 YRNEETWVDSCTECHQNSVT 346

RESULT 6
US-07-899-535A-1
; Sequence 1, Application US/07899535A
; Patent No. 5428011

; GENERAL INFORMATION:
; APPLICANT: Sheth, Anil R.
; APPLICANT: Garde, Seema
; APPLICANT: Panchal, Chandra J.
; TITLE OF INVENTION: Pharmaceutical Preparations For
; TITLE OF INVENTION: Inhibiting Tumours Associated With Prostate
; TITLE OF INVENTION: Adenocarcinoma, Stomach Cancer and Breast Cancer.
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mr. George Loud
; STREET: 2001 Jefferson Davis Highway, Suite 306
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/899,535A
; FILING DATE: 16-JUN-1992
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Loud, George A.
; REGISTRATION NUMBER: 25,814
; REFERENCE/DOCKET NUMBER: S6B-A835
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-415-0960
; TELEFAX: 703-415-0962
; TELEX: 24 8614
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 94 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-07-899-535A-1

Query Match 44.5%; Score 57; DB 1; Length 94;
Best Local Similarity 57.1%; Pred. No. 0.3;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 NETWVDSCCTCTC 17
; : | | | | | | | | | |
Db 29 NSEWQTDNCCTCTC 42

RESULT 7
US-08-726-525-2
; Sequence 2, Application US/08726525
; Patent No. 5789181
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; TITLE OF INVENTION: BINDING
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,525
; FILING DATE: 07-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,942
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A. 32,724
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: GI5258
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-726-525-2

Query Match 44.5%; Score 57; DB 1; Length 176;
Best Local Similarity 52.9%; Pred. No. 0.54;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 2 AENETWVVDSCCTCTCK 18
Db 121 ANNTKWKKDACTICECK 137

RESULT 8
US-08-487-942-2
; Sequence 2, Application US/08487942
; Patent No. 5817476
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,942
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: GI5258
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-726-036A-2

Query Match 44.5%; Score 57; DB 2; Length 176;
Best Local Similarity 52.9%; Pred. No. 0.54;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 2 AENETWVVDSCCTCTCK 18
Db 121 ANNTKWKKDACTICECK 137

RESULT 9
US-08-726-036A-2
; Sequence 2, Application US/08726036A
; Patent No. 5981482
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,036A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: GI5258
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-726-036A-2

Query Match 44.5%; Score 57; DB 2; Length 176;
Best Local Similarity 52.9%; Pred. No. 0.54;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 2 AENETWVVDSCCTCTCK 18
Db 121 ANNTKWKKDACTICECK 137

RESULT 10
US-09-083-516-2
; Sequence 2, Application US/09083516
; Patent No. 6300086
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
```


;; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
;; TITLE OF INVENTION: BINDING
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
;; STREET: 87 CambridgePark Drive
;; CITY: Cambridge
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02140
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/083,516
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/487,942
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Brown, Scott A.
;; REGISTRATION NUMBER: 32,724
;; REFERENCE/DOCKET NUMBER: GI5258
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 498-8224
;; TELEFAX: (617) 876-5851
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 176 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-09-083-516-2

Query Match 44.5%; Score 57; DB 4; Length 176;
Best Local Similarity 52.9%; Pred. No. 0.54;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 2 AENETWVVDSCCTTCTCK 18
| | | | | | | | | |
Db 121 ANNTKWKDKACTICECK 137

RESULT 11
US-08-820-170A-37
; Sequence 37, Application US/08820170A
; Patent No. 5831058
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,170A
; FILING DATE:

;; CLASSIFICATION: 536
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 293-7060
;; TELEFAX: (202) 293-7860
;; TELEX: 6491103
;; INFORMATION FOR SEQ ID NO: 37:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 816 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-820-170A-37

Query Match 39.5%; Score 50.5; DB 2; Length 816;
Best Local Similarity 47.1%; Pred. No. 18;
Matches 8; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

Qy 1 FAENETWVVDSCCTTCTC 17
: | | : | | | | |
Db 281 YREFESW-IDGCKNCTC 296

RESULT 12
US-09-055-699-37
; Sequence 37, Application US/09055699
; Patent No. 6005088
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,699
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/820,170
; FILING DATE:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-055-699-37

Query Match 39.5%; Score 50.5; DB 3; Length 816;
Best Local Similarity 47.1%; Pred. No. 18;
Matches 8; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

Qy 1 FAENETWVVDSCCTTCTC 17
: | | : | | | | |
Db 281 YREFESW-IDGCKNCTC 296

RESULT 13
US-09-273-565-37
; Sequence 37, Application US/09273565A
; Patent No. 6166190
; GENERAL INFORMATION:
; APPLICANT: FUJIWARA, TSUTOMU
; APPLICANT: WATANABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
; FILE REFERENCE: Q-53599
; CURRENT APPLICATION NUMBER: US/09/273,565A
; CURRENT FILING DATE: 1999-03-22
; EARLIER APPLICATION NUMBER: 09/055,699
; EARLIER FILING DATE: 1998-04-07
; EARLIER APPLICATION NUMBER: 08/820,170
; EARLIER FILING DATE: 1997-03-19
; EARLIER APPLICATION NUMBER: JP 63410/1996
; EARLIER FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: JP 69163/1997
; EARLIER FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 816
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-273-565-37

Query Match 39.5%; Score 50.5; DB 4; Length 816;
Best Local Similarity 47.1%; Pred. No. 18;
Matches 8; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 1 FAENETWVVDSCITCTC 17
; : : : : :
Db 281 YREFESW-IDGCKNCTC 296

RESULT 14
US-08-820-170A-34
; Sequence 34, Application US/08820170A
; Patent No. 5831058
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/820,170A
; FILING DATE:
; CLASSIFICATION: 536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 34:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 810 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-820-170A-34

Query Match 38.7%; Score 49.5; DB 2; Length 810;
Best Local Similarity 42.1%; Pred. No. 25;
Matches 8; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

QY 1 FAENETWV-VDSCTTCTCK 18
; : : : : :
Db 280 YRDQDSWVDGDHCRNCTCK 298

RESULT 15
US-09-055-699-34
; Sequence 34, Application US/09055699
; Patent No. 6005088
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/055,699
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/820,170
; FILING DATE:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 810 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-055-699-34

Query Match 38.7%; Score 49.5; DB 3; Length 810;
Best Local Similarity 42.1%; Pred. No. 25;
Matches 8; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

QY 1 FAENETWV-VDSCTTCTCK 18
; : : : : :
Db 280 YRDQDSWVDGDHCRNCTCK 298

Search completed: January 31, 2002, 09:23:56
Job time: 221 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:23:06 ; Search time 96.28 Seconds
(without alignments)
11.540 Million cell updates/sec

Title: US-09-536-087-9

Perfect score: 82

Sequence: 1 ELIGPPKTRMSAC 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	15	21 AAB19682	Human thrombospondin
2	82	100.0	1172	21 AAB19677	Human thrombospondin
3	82	100.0	1172	21 AAB00043	Human thrombospondin
4	40	48.8	406	22 AAB81955	Gamma herpesvirus
5	39	47.6	81	22 AAB35227	M tuberculosis Rv2
6	39	47.6	107	22 AAB94966	Human protein sequ
7	39	47.6	213	21 AAG13730	Arabidopsis thalia
8	39	47.6	302	20 AAY02614	Protein encoded by
9	39	47.6	302	20 AAY02611	Protein encoded by
10	39	47.6	302	20 AAY02612	Protein encoded by
11	39	47.6	302	20 AAY08458	Rat Ptx3 protein.

12	39	47.6	302	20 AAY08461	Rat Ptx3 protein.
13	39	47.6	353	21 AAY97310	Murine 11cby recep
14	38	46.3	33	21 AAB09184	Hepatitis GB virus
15	38	46.3	555	18 AAW34563	Thermotoga maritim
16	38	46.3	555	19 AAW49867	Thermotoga maritim
17	38	46.3	1964	20 AAW55557	Mus musculus notch
18	38	46.3	3164	16 AAR94346	Hepatitis GB virus
19	37	45.1	128	21 AAG09344	Arabidopsis thalia
20	37	45.1	136	21 AAG09343	Arabidopsis thalia
21	37	45.1	245	22 AAG92980	Human protein sequ
22	37	45.1	319	22 AAG91099	C glutamicum prote
23	37	45.1	319	22 AAB76694	Corynebacterium g1
24	37	45.1	336	22 AAG81765	S. epidermidis ope
25	37	45.1	336	22 AAG82753	S. epidermidis ope
26	37	45.1	426	13 AAR29863	HCV NS2-NS4 peptid
27	37	45.1	426	13 AAR29864	HCV NS2-NS4 peptid
28	37	45.1	439	14 AAR33745	XR4. Mus musculus
29	37	45.1	440	17 AAR92479	Peroxisome prolife
30	37	45.1	440	20 AAW99597	Mouse peroxisome p
31	37	45.1	828	21 AAG36819	Arabidopsis thalia
32	37	45.1	907	21 AAG36818	Arabidopsis thalia
33	37	45.1	912	21 AAG36817	Arabidopsis thalia
34	37	45.1	1188	13 AAR29870	HCV NS2-NS4 peptid
35	36.5	44.5	72	17 AAR94898	Human trophoblast ex
36	36.5	44.5	72	21 AAB23365	Human trophoblast ex
37	36.5	44.5	557	22 AAM14990	Peptide #1424 enco
38	36.5	44.5	557	22 AAM27425	Peptide #1462 enco
39	36.5	44.5	557	22 AAM02717	Peptide #1399 enco
40	36.5	44.5	674	21 AAB23358	Human trophoblast pr
41	36.5	44.5	749	17 AAR94895	Human trophoblast pr
42	36.5	44.5	749	21 AAB23357	Human trophoblast pr
43	36.5	44.5	814	22 AAB94744	Human protein sequ
44	36	43.9	40	22 AAG77276	Human colon cancer
45	36	43.9	74	20 AAY31802	Beak and feather d

ALIGNMENTS

RESULT 1
AAB19682
ID AAB19682 standard; Peptide; 15 AA.
XX AC
AA19682;
XX AC
DT 05-FEB-2001 (first entry)
XX Human thrombospondin-2 peptide 4 (aa306-320).
DE Thrombospondin-2; TSP-2; human; angiogenesis; cell proliferation;
KW melanoma; tumour; cancer; squamous cell carcinoma; antiangiogenic;
KW prostate cancer; psoriasis; rosecea dermatosis; antitumour;
KW therapy.

XX Homo sapiens.
OS
XX
XX WO200057899-A1.
PN
XX
XX 05-OCT-2000.
PD
XX
XX 24-MAR-2000; 2000WO-US07835.
PF
XX
XX 31-MAR-1999; 99US-0127221.
PR
XX
XX (GEO) GEN HOSPITAL CORP.
PA
XX
XX Detmar M, Streit M;
PI
XX
XX WPI; 2000-656131/63.
DR
XX
XX Treating a disorder characterized by unwanted cell proliferation e.g.
PT precancerous, cancerous or neoplastic cells or presence of tumor
PT preferably of skin or prostate, comprises increasing thrombospondin-2

```

PT activity -
XX Disclosure; Page 40; 73pp; English.
XX
XX The present sequence is that of peptide 4, derived from the
XX procollagen domain (amino acids 306-320) of human thrombospondin-2
XX (TSP-2, see AAB19677). The peptide is 1 of 5 synthetic peptides (see
XX AAB19679-83) used with human dermal microvascular endothelial cells
XX (HMEC) to determine the effect of TSP-2 on endothelial cell
XX migration. Peptide 4 did not significantly modify HMEC migration,
XX in contrast to peptide 7 (see AAB19683) from a type 1 repeat of
XX TSP-2, which inhibited migration. The invention is based on the
XX discovery that overexpression of TSP-2 decreases tumor size in
XX vivo, and features methods for modulating unwanted angiogenesis and
XX cell proliferation by increasing TSP-2 activity.
XX Sequence 15 AA;
SQ
Query Match 100.0%; Score 82; DB 21; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ELIGGPPKTRNMSAC 15
Db 1 eliggppktrnmsac 15
|||||
RESULT 2
AAB19677
ID AAB19677 standard; Protein; 1172 AA.
XX
XX AAB19677;
XX
XX 05-FEB-2001 (first entry)
XX
XX Human thrombospondin-2.
XX
XX Thrombospondin-2; TSP-2; human; angiogenesis; cell proliferation;
XX melanoma; tumor; cancer; squamous cell carcinoma; antiangiogenic;
XX prostate cancer; psoriasis; rosecea dermatosis; antitumor;
XX therapy.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Region 382..429 /note= "type 1 repeat"
XX Region 384..390 /note= "WSPWAEW sequence involved in antiangiogenic
XX activity"
XX Region 438..490 /note= "type 1 repeat"
XX Region 495..547 /note= "type 1 repeat"
XX
XX WO200057899-A1.
XX
XX 05-OCT-2000.
XX
XX 24-MAR-2000; 2000WO-US07835.
XX
XX 31-MAR-1999; 99US-0127221.
XX
XX (GEO ) GEN HOSPITAL CORP.
XX
XX Detmar M, Streit M;
XX PI
XX WPI; 2000-656131/63.
XX DR N-PSDB; AAA88669.
XX
XX Treating a disorder characterized by unwanted cell proliferation e.g.
XX precancerous, cancerous or neoplastic cells or presence of tumor
XX
PT preferably of skin or prostate, comprises increasing thrombospondin-2
PT activity
XX Disclosure; Fig 2; 73pp; English.
XX
XX The present sequence is that of human thrombospondin-2 (TSP-2).
XX The invention is based on the discovery that overexpression of
XX TSP-2 decreases tumor size in vivo, and features methods for
XX modulating unwanted angiogenesis and tumor growth. Treatment of
XX unwanted cell proliferation or angiogenesis involves increasing
XX TSP-2 activity. This is achieved by administering an agent which
XX increases TSP-2 activity, especially a TSP-2 polypeptide, a TSP-2
XX derived polypeptide or retro-inverso peptide, a nucleic acid
XX encoding TSP-2, an agonist of TSP-2, or an agent that increases
XX TSP-2 gene expression. The TSP-2 polypeptide may include at least
XX one type 1 repeat such as the WSPWAEW peptide (see AAB19683),
XX which is involved in the antiangiogenic activity of TSP-2. The
XX method is used to treat a disorder characterised by pre-cancerous,
XX cancerous or neoplastic cells, or the presence of a tumour, or a
XX disorder that affects epithelial tissues resulting in unwanted
XX skin cell proliferation. Such disorders include malignant
XX melanoma, prostate cancer, squamous cell carcinoma, aged skin,
XX rosecea dermatosis, psoriasis, and skin damage caused by
XX photoradiation (all claimed). Evaluating the presence of TSP-2
XX nucleic acid or protein is useful for diagnosing a subject at risk
XX of unwanted cell proliferation or angiogenesis. Methods are also
XX provided of identifying compounds that modulate TSP-2 activity.
XX
XX Sequence 1172 AA;
SQ
Query Match 100.0%; Score 82; DB 21; Length 1172;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ELIGGPPKTRNMSAC 15
Db 306 eliggppktrnmsac 320
|||||
RESULT 3
AAB00043
ID AAB00043 standard; Protein; 1172 AA.
XX
XX AAB00043;
XX
XX 08-NOV-2000 (first entry)
XX
XX Human thrombospondon-2 (TSP-2).
XX
XX TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
XX thrombospondin; angiogenesis; tumor; treatment; cancer;
XX arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
XX glaucoma.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Region 381..436 /label= Type 1 repeat region
XX Region 437..493 /label= Type 1 repeat region
XX Region 494..550 /label= Type 1 repeat region
XX
XX WO200044908-A2.
XX
XX 03-AUG-2000.
XX
XX 01-FEB-2000; 2000WO-US02482.
XX
XX 01-FEB-1999; 99US-0118053.
XX

```

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI

XX Lawler JW;

XX WPI; 2000-514823/46.

XX Nucleic acids encoding chimeric proteins such as cartilage oligomeric
PT matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for
PT inhibiting angiogenesis and treating diseases such as cancer

XX Disclosure; Fig 2; 40pp; English.

XX New nucleic acids are described which encode a protein comprising
CC the second and third type 1 repeats of human TSP (thrombospondin)-1,
CC but not the TGF (transforming growth factor)-beta activation region
CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing
CC the second and third type-1 repeats and the COMP (cartilage
CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was
CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1
CC caused inhibition of the growth of tumours in mice models.

XX Thus the nucleic acids and proteins may be useful for treating
CC angiogenesis related diseases such as cancer (by reducing the rate of
CC growth and size of tumours), arthritis, psoriasis, diabetic
CC retinopathy, corneal graft rejection, and glaucoma. They may also be
CC used for treating human immunodeficiency virus (HIV) infection.
CC Anti-angiogenic therapy has little toxicity, does not require the
CC therapeutic agent to enter tumour cells or cross the blood-brain
CC barrier, controls tumour growth independently of growth of
CC tumour cell heterogeneity, and does not induce drug resistance.

XX Sequence 1172 AA;

Query Match 100.0%; Score 82; DB 21; Length 1172;

Best Local Similarity 100.0%; Pred. No. 1.4e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ELIGGPKTRNMSAC 15

DB 306 eliggppktrnmsac 320

RESULT 4

AAB81955

ID AAB81955 standard; Protein; 406 AA.

XX AAB81955;

XX 02-JUL-2001 (first entry)

XX Gamma herpesvirus-68 M3 protein.

XX GammaHV68; M3 protein; human; inflammatory disorder; autoimmune disease;
KW immunomodulatory; cancer; neurodegenerative disease.

XX Gammaherpesvirus-68.

XX WO200121204-A1.

XX 29-MAR-2001.

XX 20-SEP-2000; 2000WO-US25871.

XX 20-SEP-1999; 99US-0154799.

XX (ROBA-) ROBERTS RES INST JOHN P.
PA (UNIW) UNIV WASHINGTON.

XX McFadden G, Speck S, Virgin H, Lucas A;

XX WPI; 2001-316079/33.

XX N-PSDB; AAF85695, AAF85696.

PT Treating inflammatory and autoimmune disorders, e.g. rheumatoid
PT arthritis, transplant rejection, restenosis, asthma, allergies, or
PT systemic lupus erythematosus, comprises administering an M3 polypeptide

XX Claim 2; Fig 3; 67pp; English.

XX The present invention describes a method of immunomodulation in a mammal,
CC involving the administration of M3 protein from gammaherpesvirus-68
CC (gammaHV68). This is useful in the treatment of cancer, autoimmune and
CC inflammatory diseases, including neurodegeneration, acute inflammation,
CC rheumatoid arthritis, transplant rejection, restenosis, asthma,
CC allergies, pollinosis, systemic lupus erythematosus, nephrotic syndrome
CC dermatitis, multiple sclerosis, myasthenia gravis, diabetes,
CC glomerulonephritis, Hashimoto's thyroiditis, autoimmune haemolytic
CC anaemia or thrombocytopaenic purpura, microbial infection, cirrhosis,
CC endotoxaemia, atherosclerosis, AIDS and nervous system diseases. The
CC present sequence is the M3 protein.

XX Sequence 406 AA;

Query Match 48.8%; Score 40; DB 22; Length 406;

Best Local Similarity 50.0%; Pred. No. 77;

Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 GGPPKTRNMSAC 15

DB 275 ggppqakmmtc 286

RESULT 5

AAB35227

ID AAB35227 standard; Protein; 81 AA.

XX AAB35227;

XX 24-APR-2001 (first entry)

XX M tuberculosis Rv2654c protein.

XX Tuberculosis; TB; vaccine; esat-6 gene family; Rv0287; Rv1036c;
KW Rv1037c; Rv2346c; Rv2348c; Rv2653c; Rv2654c; Rv3020c; Rv3444c;
KW Rv3445c; Rv3890c; Rv3891c; Rv3904c; Rv3905c.

XX Mycobacterium tuberculosis.

XX WO200104151-A2.

XX 18-JAN-2001.

XX 13-JUL-2000; 2000WO-DK00398.

XX 13-JUL-1999; 99DK-0001020.

XX 15-JUL-1999; 99US-0144011.

XX (STAT-) STATENS SERUM INST.

XX Andersen P, Skjot R;

XX WPI; 2001-091923/10.

XX N-PSDB; AAF24417.

XX New polypeptide encoded by a member of the esat-6-gene family for
PT immunizing against and diagnosis of tuberculosis -

XX Claim 2; Page 72-73; 80pp; English.

XX The present invention provides the protein and coding sequences for
CC members of the esat-6 gene family from Mycobacterium tuberculosis. These
CC proteins include Rv0287, Rv1036c, Rv1037c, Rv2346c, Rv2348c, Rv2653c,
CC Rv2654c, Rv3020c, Rv3444c, Rv3445c, Rv3890c, Rv3891c, Rv3904c and

CC Rv3905c. These can be used to produce vaccines against, and in the
CC diagnosis of, tuberculosis (TB) infection. The present sequence is one of
CC the proteins of the invention.

XX
SQ Sequence 81 AA;

Query Match	47.6%	Score 39;	DB 22;	Length 81;
Best Local Similarity	50.0%	Pred. No. 22;		
Matches	7;	Conservative	3;	Mismatches 4;
				Indels 0;
				Gaps 0;

Qy 1 ELIGGPPKTRNMSA 14
 ||:|||| : :|
Db 17 elvggpppveasaaa 30

RESULT 6
AAB94966
ID AAB94966 standard; Protein; 107 AA.

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:16528.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999: 99JP-0248036.

PR 27 AUG 1953; 330F-0300233
PR 11-JAN-2000; 2000JP-0118776.

PR 09-JUN-2000: 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T. Isodai T. Nish

XX
TITIST S, sugryana T,
XX

XX
WFL; 2001-318/49/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -

PS Claim 8: SEO ID 16528: 2537pp + CD ROM: English

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length

PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143342.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.

PR 29-OCT-1999; 99US-0162142.

Query Match 47.6%; Score 39; DB 21; Length 213;
Best Local Similarity 60.0%; Pred. No. 59;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 IGGPPKTRNM 12
: ||| ||: |
Db 9 lggptktgql 18

RESULT 8

AA02614
ID AAY02614 standard; Protein; 302 AA.

XX AC AAY02614;

XX DT 21-JUL-1999 (first entry)

XX DE Protein encoded by the human mutant Pitx3 gene.

XX KW Pitx3; homeobox domain protein; lens development; lens disorder;
XX KW cataract; detection; ocular disease; ASMD; Peter's anomaly;
XX KW anterior segment mesenchymal dysgenesis; mutant.

XX OS Homo sapiens.

XX PN WO9921996-A1.

XX PD 06-MAY-1999.

XX PF 26-OCT-1998; 98WO-US22689.

XX PR 24-OCT-1997; 97US-0957351.

XX PA (IOWA) UNIV IOWA RES FOUND.

XX PI Murray JC, Semina EV;

XX DR WPI; 1999-312965/26.
XX DR N-PSDB; AAX59565.

XX PT Pitx3, homeobox protein, and related nucleic acid sequences

XX PS Claim 13; Fig 7; 128pp; English.

XX CC The present sequence represents a mutant human Pitx3 protein, in which
XX CC amino acid 13 is different from the wild type protein. Pitx3 proteins
XX CC are homeobox domain proteins, which are involved in the development
XX CC of the lens and contribute to diseases and disorders of the lens,
XX CC such as cataracts. The Pitx3 nucleic acids (e.g. antisense sequences,
XX CC ribozymes and triplex nucleic acids), probes derived from them and
XX CC polypeptides, are useful in claimed methods to detect an ocular
XX CC disease, especially of the lens, e.g. cataract formation. Specific
XX CC conditions that can be detected and treated are Anterior Segment
XX CC Mesenchymal Dysgenesis (ASMD) and Peter's anomaly.

XX SQ Sequence 302 AA;

Query Match 47.6%; Score 39; DB 20; Length 302;
Best Local Similarity 46.2%; Pred. No. 85;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 IGGPPKTRNMSAC 15
: ||| ||: |
Db 282 vggpppaanlspc 294

RESULT 9

AA02611
ID AAY02611 standard; Protein; 302 AA.

XX

AC AAY02611;

XX DT 21-JUL-1999 (first entry)

XX DE Protein encoded by the mouse Pitx3 gene.

XX KW Pitx3; homeobox domain protein; lens development; lens disorder;
XX KW cataract; detection; ocular disease; ASMD; Peter's anomaly;
XX KW anterior segment mesenchymal dysgenesis.

XX OS Mus sp.

XX PN WO9921996-A1.

XX PD 06-MAY-1999.

XX PF 26-OCT-1998; 98WO-US22689.

XX PR 24-OCT-1997; 97US-0957351.

XX PA (IOWA) UNIV IOWA RES FOUND.

XX PI Murray JC, Semina EV;

XX DR WPI; 1999-312965/26.
XX DR N-PSDB; AAX59545.

XX PT Pitx3, homeobox protein, and related nucleic acid sequences

XX PS Claim 13; Fig 1; 128pp; English.

XX CC The present sequence is encoded by the mouse Pitx3 gene. Pitx3 proteins
XX CC are homeobox domain proteins, which are involved in the development
XX CC of the lens and contribute to diseases and disorders of the lens,
XX CC such as cataracts. The Pitx3 nucleic acids (e.g. antisense sequences,
XX CC ribozymes and triplex nucleic acids), probes derived from them and
XX CC polypeptides, are useful in claimed methods to detect an ocular
XX CC disease, especially of the lens, e.g. cataract formation. Specific
XX CC conditions that can be detected and treated are Anterior Segment
XX CC Mesenchymal Dysgenesis (ASMD) and Peter's anomaly.

XX SQ Sequence 302 AA;

Query Match 47.6%; Score 39; DB 20; Length 302;
Best Local Similarity 46.2%; Pred. No. 85;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 IGGPPKTRNMSAC 15
: ||| ||: |
Db 282 vggpppaanlspc 294

RESULT 10

AA02612
ID AAY02612 standard; Protein; 302 AA.

XX AC AAY02612;

XX DT 21-JUL-1999 (first entry)

XX DE Protein encoded by the human Pitx3 gene.

XX KW Pitx3; homeobox domain protein; lens development; lens disorder;
XX KW cataract; detection; ocular disease; ASMD; Peter's anomaly;
XX KW anterior segment mesenchymal dysgenesis.

XX OS Homo sapiens.

XX PN WO9921996-A1.

XX PD 06-MAY-1999.

XX

```

PF 26-OCT-1998; 98WO-US22689.
XX
XX 24-OCT-1997; 97US-0957351.
XX
XX (IOWA ) UNIV IOWA RES FOUND.
XX
XX Murray JC, Semina EV;
PI
XX WPI; 1999-312965/26.
XX
XX N-PSDB; AAX59546.
XX
XX Pitx3, homeobox protein, and related nucleic acid sequences
PT
XX
XX Claim 13; Fig 3; 128pp; English.
PS
XX
XX The present sequence is encoded by the human Pitx3 gene. Pitx3 proteins
CC are homeobox domain proteins, which are involved in the development
CC of the lens and contribute to diseases and disorders of the lens,
CC such as cataracts. The Pitx3 nucleic acids (e.g. antisense sequences,
CC ribozymes and triplex nucleic acids), probes derived from them and
CC polypeptides, are useful in claimed methods to detect an ocular
CC disease, especially of the lens, e.g. cataract formation. Specific
CC conditions that can be detected and treated are Anterior Segment
CC Mesenchymal Dysgenesis (ASMD) and Peter's anomaly.
XX
XX
XX Sequence 302 AA;
SQ
Query Match 47.6%; Score 39; DB 20; Length 302;
Best Local Similarity 46.2%; Pred. No. 85;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
Qy 3 IGGPPKTRNMSAC 15
Db : ||| | : | |
282 vhgpppaanlspc 294
RESULT 11
RAY08458
ID AAY08458 standard; Protein; 302 AA.
XX
XX AAY08458;
XX
XX 27-JUL-1999 (first entry)
XX
XX Rat Ptx3 protein.
XX
XX Rat; Ptx3; mesencephalic dopaminergic neuron; disorder; neurological;
KW psychiatric; medication; Parkinson's disease; tardive dyskinesia;
KW manic depression; schizophrenia.
XX
XX Rattus sp.
OS
XX
XX WO9924572-A1.
XX
XX 20-MAY-1999.
XX
XX 11-NOV-1998; 98WO-NL00652.
XX
XX 11-NOV-1997; 97EP-0203511.
XX
XX (UYUT-) RIJKSUNIV UTRECHT.
XX
XX Burbach JPH, Smidt MP;
PI
XX WPI; 1999-327404/27.
XX
XX Gene with restricted expression in mesencephalic dopaminergic
DR neurons
XX
XX Disclosure; Fig 1; 57pp; English.
XX
XX This invention describes a novel gene with restricted expression in
CC

```

RESULT 13
 ID AAY97310 standard; Protein; 353 AA.
 XX AC AAY97310;
 XX AC AAY97310;
 DT 03-JAN-2001 (first entry)
 XX DE Murine llcby receptor.
 XX DE
 XX llcby receptor; agonist; antagonist; high throughput screening;
 KW narrowwell method; bacteria; fungi; protozoa; virus; HIV; cancer;
 KW diabetes; obesity; anorexia; bulimia; asthma; Parkinson's disease;
 KW acute heart failure; hypotension; hypertension; urinary retention;
 KW osteoporosis; angina pectoris; myocardial infarction; stroke;
 KW ulcer; allergy; migraine; vomiting; neurological disorder;
 KW schizophrenia; manic depression; delirium; dementia;
 KW mental retardation; mouse.
 XX OS
 XX Mus musculus.
 XX WO2000049170-A1.
 XX PD 24-AUG-2000.
 XX PF 20-JAN-2000; 2000WO-US01358.
 XX PR 18-FEB-1999; 99US-0252441.
 XX PA (SMIK) SMITHKLINE BEECHAM CORP.
 XX PI Bergsma D, Ellis CE;
 XX WPI; 2000-549280/50.
 XX N-PSDB; AAA53886.
 PT Isolated murine llcby polynucleotides and polypeptides, useful for
 PT identifying llcby receptor agonists and antagonists that may be
 PT potentially useful in therapy
 XX Claim 1; Page 28; 33pp; English.
 XX The murine llcby receptor and nucleotides encoding it may be used to
 CC identify llcby receptor agonists and antagonists that may be
 CC potentially useful in therapy. The receptor polypeptide may be
 CC employed in conventional low capacity screening methods and also in
 CC high-throughput screening (HTS) formats. Such HTS formats include not
 CC only the well-established use of 96 and 384-well microtiter plates
 CC but also emerging methods such as the nanowell method. Antagonists to
 CC the llcby receptor are thought to have application in treating a
 CC variety of human diseases caused by bacteria, fungi, protozoa and
 CC viruses including HIV, cancer, diabetes, obesity, anorexia, bulimia,
 CC asthma, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC myocardial infarction, stroke, ulcers, allergies, migraine,
 CC vomiting, neurological disorders, schizophrenia, manic depression,
 CC delirium, dementia and severe mental retardation.
 XX Sequence 353 AA;

Query Match 47.6%; Score 39; DB 21; Length 353;
 Best Local Similarity 58.3%; Pred. No. 1e+02;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 2 LIGGPPKTRNMS 13
 DB 26 lagppptrsvs 37
 | | | | | | | | | |
 RESULT 14

AAB09184
 ID AAB09184 standard; Protein; 33 AA.
 XX AC AAB09184;
 XX DT 30-AUG-2000 (first entry)
 XX DE Hepatitis GB virus protein sequence SEQ ID NO:306.
 XX KW Hepatitis GB virus; HGBV; diagnosis; therapeutic; immunogenic;
 KW infection; detection; characterisation; hepatitis.
 XX OS
 XX Hepatitis GB virus.
 XX US6051374-A.
 XX PD 18-APR-2000.
 XX PF 07-JUN-1995; 95US-0488445.
 XX PR 14-FEB-1994; 94US-0196030.
 XX PR 13-MAY-1994; 94US-0242654.
 XX PR 29-JUL-1994; 94US-0283314.
 XX PR 23-NOV-1994; 94US-0344185.
 XX PR 23-NOV-1994; 94US-0344190.
 XX PR 30-JAN-1995; 95US-0377557.
 XX PA (ABBO) ABBOTT LAB.
 XX Dawson GJ, Leary TP, Muerhoff AS, Pilot-Matias TJ, Buijk SL;
 PI Mushahwar IK, Simons JN, Desai SM, Erker JC, Schlauder GG;
 XX WPI; 2000-338307/29.
 XX Detecting target hepatitis GB virus nucleic acid in a test sample
 PT suspected of containing HGBV comprises reacting the test sample the
 PT HGBV polynucleotide probe and detecting the complex that contains
 PT target HGBV -
 XX Example 9; Column 357-358; 369pp; English.
 XX The present invention describe a method for detecting target
 CC hepatitis GB virus (HGBV) nucleic acid (THN) in a test sample (T)
 CC suspected of containing HGBV. The method involves reacting (T) with a
 CC HGBV polynucleotide probe (I) containing 15 contiguous nucleotides, and
 CC which selectively hybridises to the HGBV genome or its full complement,
 CC and detecting the complex that contains THN, indicating the presence of
 CC target HGBV. The method is used for detecting target HGBV nucleic acid
 CC in the test sample suspected of containing HGBV and for characterisation
 CC of newly ascertained etiological agent of non-A, non-B, non-C, non-D and
 CC non-E hepatitis causing agents collectively termed as hepatitis GB
 CC virus. AAA55270 to AAA55489 and AAB08985 to AAB09480 represent nucleotide
 CC and protein sequences used in the exemplification of the present
 CC invention.
 XX Sequence 33 AA;
 Query Match 46.3%; Score 38; DB 21; Length 33;
 Best Local Similarity 77.8%; Pred. No. 13;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 3 IGGPPKTRN 11
 DB 3 igpppcarn 11
 | | | | | | | | | |
 RESULT 15
 AAW34563
 ID AAW34563 standard; Protein; 555 AA.
 XX AC AAW34563;
 XX

12-MAR-1998 (first entry)
Thermotoga maritima alpha-galactosidase.
Glycosidase; thermostable; textile; food processing; pharmaceutical;
detergent; baking; industry; Thermococcus; Staphylothermus;
Pyrococcus; glucose; soluble oligosaccharide; alpha-galactosidase.
Thermotoga maritima.
WO9725417-A1.
17-JUL-1997.
10-JAN-1997; 97WO-US000092.
13-SEP-1996; 96US-0712612.
11-JAN-1996; 96US-0583787.
(RECO-) RECOMBINANT BIOCATALYSIS INC.
Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
WPI; 1997-372858/34.
N-PSDB; AAT93687.
New thermostable glycosidase(s) - from Thermococcus, Staphylothermus
and Pyrococcus, used in the textile, food processing,
pharmaceutical, detergent and baking industries
Claim 4; Fig 10; 82pp; English.
The present sequence represents alpha-galactosidase isolated from
Thermotoga maritima. The enzyme or its encoding nucleic acid sequence is
used for generating glucose from soluble oligosaccharides. The enzyme
can be used in the food processing, pharmaceutical, textile, detergent
and baking industries. The enzyme is also used to treat lactose
intolerance, as a diagnostic reporter molecule, in corn wet milling or
in the fruit juice industry. The enzymes can be used to hydrolyse guar
gum to remove non-reducing terminal mannose residues. The nucleic acids
encoding the enzyme may be used to generate probes to identify similar
sequences.
Sequence 555 AA;
Query Match 46.3%; Score 38; DB 18; Length 555;
Best Local Similarity 50.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 1 ELIGGPPKTRNM 12
||| | : | :
Db 482 ellggrprvqni 493
Search completed: January 31, 2002, 09:23:06
Job time: 196 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:36:16 ; Search time 90.11 Seconds
(without alignments)
24.349 Million cell updates/sec

Title: US-09-536-087-9
Perfect score: 82
Sequence: 1 ELIGGPPKTRNMSAC 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	52.4	351	10 Q9SST9	Q9sst9 scaparia du
2	41	50.0	140	3 Q9P8A5	Q9p8a5 agaricus bi
3	41	50.0	439	10 Q9FWF7	Q9fwf7 oryza sativ
4	40.5	49.4	784	5 Q9GYW0	Q9gyw0 caenorhabdi
5	40.5	49.4	958	5 Q62397	Q62397 caenorhabdi
6	40	48.8	406	12 Q41925	Q41925 murid herpe
7	40	48.8	514	10 Q9LX52	Q9lx52 arabidopsis
8	40	48.8	665	2 Q07314	Q07314 rhodobacter
9	39	47.6	81	2 P71951	P71951 mycobacteri
10	39	47.6	219	12 Q71119	Q71119 trichoplusi
11	39	47.6	220	12 Q9PY04	Q9pyu4 xestia c-ni
12	39	47.6	249	11 Q9ER80	Q9er80 mus musculu
13	39	47.6	249	11 Q9D3D6	Q9d3d6 mus musculu
14	39	47.6	651	10 Q9C9Q4	Q9c9q4 arabidopsis
15	39	47.6	815	5 Q9V9W8	Q9v9w8 drosophila
16	39	47.6	1329	5 Q9W283	Q9w283 drosophila
17	38.5	47.0	391	2 Q69953	Q69953 streptomyce
18	38	46.3	149	4 Q13038	Q13038 homo sapien
19	38	46.3	207	2 Q54312	Q54312 streptomyce

20	38	46.3	260	2 Q9ILL3	Q9ill3 pseudomonas
21	38	46.3	363	5 Q21427	Q21427 caenorhabdi
22	38	46.3	405	2 Q9FCW7	Q9fcm7 streptomyce
23	38	46.3	470	1 Q27166	Q27166 methanobact
24	38	46.3	552	2 Q33835	Q33835 thermotoga
25	38	46.3	556	10 Q9FZ82	Q9fz82 arabidopsis
26	38	46.3	1024	5 Q09570	Q09570 caenorhabdi
27	38	46.3	1964	11 Q35442	Q35442 mus musculu
28	37.5	45.7	413	10 Q9FKK6	Q9fkk6 arabidopsis
29	37.5	45.7	568	10 Q04585	Q04585 arabidopsis
30	37	45.1	128	10 Q23020	Q23020 arabidopsis
31	37	45.1	159	10 Q9M117	Q9m117 arabidopsis
32	37	45.1	242	12 P89036	P89036 panicum mos
33	37	45.1	245	4 Q9NVP6	Q9nvp6 homo sapien
34	37	45.1	259	11 Q9D5P1	Q9d5p1 mus musculu
35	37	45.1	271	11 Q9D4L6	Q9d4l6 mus musculu
36	37	45.1	295	2 Q9HWY4	Q9hwy4 pseudomonas
37	37	45.1	298	10 Q40839	Q40839 picea glauc
38	37	45.1	343	4 Q92519	Q92519 homo sapien
39	37	45.1	343	6 Q28283	Q28283 canis famil
40	37	45.1	344	10 Q9LJY2	Q9ljy2 arabidopsis
41	37	45.1	348	1 Q27144	Q27144 methanobact
42	37	45.1	367	2 Q9F924	Q9f924 pasteurella
43	37	45.1	367	2 Q9CK74	Q9ck74 pasteurella
44	37	45.1	368	10 Q24181	Q24181 oryza sativ
45	37	45.1	403	2 Q9JZ35	Q9jz35 neisseria m

ALIGNMENTS

RESULT 1					
Q9SST9					
ID Q9SST9	PRELIMINARY;	PRT;	351 AA.		
AC Q9SST9;					
DT 01-MAY-2000 (TREMBLrel. 13, Created)					
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)					
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)					
DE GERANYLGERANYL PYROPHOSPHATE SYNTHASE.					
OS Scaparia dulcis.					
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;					
OC Asteridae; euasterids I; Lamiales; Veronicaceae; Scaparia.					
OX NCBI_TaxID=107240;					
RN [1]					
RP SEQUENCE FROM N.A.					
RA Kojima N., Suh D., Sankawa U.;					
RT "Scaparia dulcis GPP synthase.";					
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.					
DR EMBL; AB034250; BAA86285.1; -					
DR InterPro; IPR000092; Polyphenyl_synt.					
DR Pfam; PF00346; polyphenyl_synt; 1.					
DR PROSITE; PS00723; POLYPHENYL_SYNTHET_1; 1.					
DR PROSITE; PS00444; POLYPHENYL_SYNTHET_2; 1.					
SQ SEQUENCE 351 AA; 38342 MW; 247570E1CD863D5C CRC64;					
Query Match	52.4%;	Score 43;	DB 10;	Length 351;	
Best Local Similarity	53.3%;	Pred. NO. 16;			
Matches	8;	Conservative	2;	Mismatches	5;
				Indels	0;
				Gaps	0;
QY 1 ELIGGPPKTRNMSAC 15					
Db 112 ELVGHPSTMPAAC 126					
	:				
RESULT 2					
Q9P8A5					
ID Q9P8A5	PRELIMINARY;	PRT;	140 AA.		
AC Q9P8A5;					
DT 01-OCT-2000 (TREMBLrel. 15, Created)					
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)					
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)					

```

DE HYPOTHETICAL 14.7 KDA PROTEIN (FRAGMENT).
GN DAG11.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Hymenomycetidae;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HORST U3;
RA Eastwood D.C., Kingsnorth C.S., Jones H., Burton K.S.;
RT "Development of Agaricus bisporus sporophores following tissue
RT detachment is maintained by the expression of stress tolerance and
RT nutritional genes.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ271703; CAB85696.1; -
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 140
SQ SEQUENCE 140 AA; 14653 MW; 182F2B69EC587285 CRC64;

Query Match 50.0%; Score 41; DB 3; Length 140;
Best Local Similarity 61.5%; Pred. No. 15;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 LIGGPPKTRNMSA 14
   ||| ||| ||| |||
Db 20 LIGSPPDQANLSA 32

RESULT 3
Q9FWF7 PRELIMINARY; PRT; 439 AA.
AC Q9FWF7
DT 01-AUG-1998 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE HYPOTHETICAL 48.6 KDA PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,
RA Hsiao J., Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OSJNB001511 genomic sequence.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AC051633; AAG13609.1; -
KW Hypothetical protein.
SQ SEQUENCE 439 AA; 48565 MW; 31A5481D7DA12B2C CRC64;

Query Match 50.0%; Score 41; DB 10; Length 439;
Best Local Similarity 53.8%; Pred. No. 45;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 ELIGGPPKTRNMS 13
   :|: ||| |||:|
Db 258 DLVDAPPGRNVS 270

RESULT 4
Q9GYW0 PRELIMINARY; PRT; 784 AA.
AC Q9GYW0
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

```

```

DE AUXILIN.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Greener T., Grant B., Zhang Y., Wu X., Green L.E., Hirsh D.,
RA Eisenberg E.;
RT "Caenorhabditis elegans auxilin is an essential protein required for
RT clathrin-mediated endocytosis in vivo.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF293972; AAG02478.1; -
DR InterPro; IPR001623; DnaJ_N.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
SQ SEQUENCE 784 AA; 86529 MW; 6573C3337CB34296 CRC64;

Query Match 49.4%; Score 40.5; DB 5; Length 784;
Best Local Similarity 40.0%; Pred. No. 97;
Matches 10; Conservative 2; Mismatches 2; Indels 11; Gaps 1;

QY 1 ELIGG-----PPKTRNMSA 14
   :|: ||| ||| ||| |||
Db 563 DLGGFSPMKPQTTPPTNPNPSA 587

RESULT 5
O62397 PRELIMINARY; PRT; 958 AA.
AC O62397
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE W07A8.3 PROTEIN.
GN W07A8.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Basham V.;
RA Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z82075; CAB04929.1; -
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
DR SMART; SM00271; DnaJ; 1.
SQ SEQUENCE 958 AA; 107100 MW; 0F8B868A9D88A6AC CRC64;

Query Match 49.4%; Score 40.5; DB 5; Length 958;
Best Local Similarity 40.0%; Pred. No. 1.2e+02;
Matches 10; Conservative 2; Mismatches 2; Indels 11; Gaps 1;

QY 1 ELIGG-----PPKTRNMSA 14

```



```

Db 700 DLGGFSPMKPQTSTPPNTPNSA 724
:|||||
RESULT 6
O41925 PRELIMINARY; PRT; 406 AA.
ID O41925
AC O41925
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE HYPOTHETICAL 44.2 KDA PROTEIN.
GN GAMMAHV.M3.
OS murid herpesvirus 4.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae.
OX NCBI_TaxID=33708;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WUMS, AND G2.4;
RX MEDLINE=97366649; PubMed=9223479;
RA Virgin H.W. IV, Latreille P., Wamsley P., Hallsworth K., Weck K.E.,
RA Dai Canto A.J., Speck S.H.;
RT "Complete sequence and genomic analysis of murine gammaherpesvirus
68."
RL J. Virol. 71:5894-5904(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WUMS;
RX MEDLINE=99214403; PubMed=10196360;
RA van Berkel V., Preiter K., Virgin H.W. IV, Speck S.H.;
RT "Identification and initial characterization of the murine
gammaherpesvirus 68 gene M3, encoding an abundantly secreted
protein."
RL J. Virol. 73:4524-4529(1999).
RN [4]
RP SEQUENCE OF 338-406 FROM N.A.
RC STRAIN=G2.4;
RX MEDLINE=97368431; PubMed=9225045;
RA Bowden R.J., Simas J.P., Davis A.J., Efsthliou S.;
RT "Murine gammaherpesvirus 68 encodes tRNA-like sequences which are
expressed during latency."
RL J. Gen. Virol. 78:1675-1687(1997).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=G2.4;
RA Milligan S., Efsthliou S., Stewart J.P., Nash A.A., Davison A.J.;
RT "Genetic content of murine gammaherpesviruses."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; U97553; AAB66428.1; -
DR EMBL; AF127083; AAD29402.1; -
DR EMBL; AF105037; AAF19271.1; -
KW Hypothetical protein.
SQ SEQUENCE 406 AA; 44231 MW; 7C8B26A064E20DB2 CRC64;

Query Match 48.8%; Score 40; DB 12; Length 406;
Best Local Similarity 50.0%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 GGPPKTRNMSAC 15
|||||
Db 275 GGPPQAKMNTC 286

RESULT 7
O41925 PRELIMINARY; PRT; 514 AA.
ID O41925

```

```

AC O9LX52
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE HYPOTHETICAL 58.4 KDA PROTEIN.
GN F25L23_50.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL356014; CAB91589.1; -
DR InterPro; IPR001810; F-box.
DR PROSITE; PS00181; F-box; 1.
KW Hypothetical protein.
SQ SEQUENCE 514 AA; 58436 MW; 01E191CD4F2748C7 CRC64;

Query Match 48.8%; Score 40; DB 10; Length 514;
Best Local Similarity 58.3%; Pred. No. 79;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 GGPPKTRNMSAC 15
|||||
Db 179 GGPPFAKLISAC 190

RESULT 8
O07314 PRELIMINARY; PRT; 665 AA.
ID O07314
AC O07314
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE NUOG (NADH:UBIQUINONE OXIDOREDUCTASE COMPLEX I SUBUNIT).
GN NUOG
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bl0;
RA Dupuis A., Issartel J.P.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-101 FROM N.A.
RX MEDLINE=97363217; PubMed=9219542;
RA Herter S.M., Schiltz E., Drews G.;
RT "Protein and gene structure of the NADH-binding fragment of
Rhodobacter capsulatus NADH:ubiquinone oxidoreductase."
RL Eur. J. Biochem. 246:800-808(1997).
DR EMBL; AF029365; AAC24995.1; -
DR EMBL; Y09884; CAA71017.1; -
DR InterPro; IPR000283; Complex1_75K.
DR InterPro; IPR001041; Ferredoxin.
DR InterPro; IPR001467; Molybdopterin.
DR Pfam; PF00111; fer2; 1.
DR Pfam; PF00384; molybdopterin; 1.
DR PROSITE; PS00641; COMPLEX1_75K_1; 1.
DR PROSITE; PS00642; COMPLEX1_75K_2; 1.
DR PROSITE; PS00643; COMPLEX1_75K_3; 1.
KW Iron-sulfur; Ubiquinone.
SQ SEQUENCE 665 AA; 71295 MW; 6B220C77C6D34428 CRC64;

```

Query Match 48.8%; Score 40; DB 2; Length 665;
 Best Local Similarity 75.0%; Pred. No. 1e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ELIGPPK 8
 I:|||||
 Db 54 EVGGPPK 61

RESULT 9
 P71951 PRELIMINARY; PRT; 81 AA.
 AC P71951;
 DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
 DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
 DE HYPOTHETICAL 7.7 KDA ALANINE-RICH PROTEIN CY441.23C PRECURSOR.
 GN RV2654C OR MTCY441.23C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RL Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A.;
 Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO THE TYPE-I AFP FAMILY. TYPE 1 AFP ARE ALANINE-RICH,
 CC AMPHIPHILIC AND ALPHA-HELICAL.
 DR EMBL; Z80225; CAB02330.1; -;
 KW Tuberculin; RV2654C; -;
 DR Hypothetical protein; Signal; Complete proteome.
 FT SIGNAL 1 38 POTENTIAL.
 FT CHAIN 39 81 HYPOTHETICAL ALANINE-RICH PROTEIN
 ET CY441.23C.
 SQ SEQUENCE 81 AA; 7698 MW; 86E5882CB1057779 CRC64;

Query Match 47.6%; Score 39; DB 2; Length 81;
 Best Local Similarity 50.0%; Pred. No. 19;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ELIGPPKTRNSA 14
 I:||||| : :
 Db 17 ELVGGPPVEASAAA 30

RESULT 10
 O71119 PRELIMINARY; PRT; 219 AA.
 AC O71119;
 DT 01-AUG-1998 (TRENBLrel. 07, Created)
 DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE P24.1.
 OS Trichoplusia ni granulosis virus (TnGV) (Trichoplusia ni
 OS granulovirus).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.
 OX NCBI_TaxID=10462;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98264509; PubMed=9603347;
 RA Bideshi D.K., Hice R.H., Ge B., Federici B.A.;
 RT "Molecular characterization and expression of the Trichoplusia ni
 RT granulovirus helicase gene."
 RL J. Gen. Virol. 79:1309-1319(1998).
 DR EMBL; AF032994; AAC40850.1; -;
 SQ SEQUENCE 219 AA; 24129 MW; E61AC3E78C359C2D CRC64;

Query Match 47.6%; Score 39; DB 12; Length 219;

Best Local Similarity 70.0%; Pred. No. 51;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 IGGPPKTRNM 12
 ||| || :||
 Db 55 IGGHPKVKNM 64

RESULT 11
 Q9PYU4 PRELIMINARY; PRT; 220 AA.
 AC Q9PYU4;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE ORF99.
 GN ORF99.
 OS Xestia c-nigrum granulosis virus (XnGV) (Xestia c-nigrum
 OS granulovirus).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.
 OX NCBI_TaxID=51677;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99434230; PubMed=10502508;
 RA Hayakawa T., Ko R., Okano K., Seong S.I., Goto C., Maeda S.;
 "Sequence analysis of the Xestia c-nigrum granulovirus genome."
 RL Virology 262:277-297(1999).
 DR EMBL; AF162221; AAF05213.1; -;
 SQ SEQUENCE 220 AA; 24232 MW; BB21450820841459 CRC64;

Query Match 47.6%; Score 39; DB 12; Length 220;
 Best Local Similarity 70.0%; Pred. No. 51;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 IGGPPKTRNM 12
 ||| || :||
 Db 55 IGGHPKVKNM 64

RESULT 12
 Q9ER80 PRELIMINARY; PRT; 249 AA.
 AC Q9ER80;
 DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE 28KD INTERFERON ALPHA RESPONSIVE PROTEIN (5830458K16RIK PROTEIN).
 GN IFRG28 OR 5830458K16RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DBA/2; TISSUE=ORO-NASAL COMPLEX;
 RA Meritet J.F., Dron M., Tovey M.;
 RT "Characterization of ifrg15 and ifrg28, two newly identified
 RT interferon responsive gene."
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CS7BL/6J; TISSUE=PANCREAS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AJ251364; CAC13976.1; -;
 DR EMBL: AK007477; BAB25057.1; -;
 DR MGD: MGI:1915025; 5830458K16Rik.
 SQ SEQUENCE 249 AA; 28391 MW; 2C79B36ED6F042D4 CRC64;

Query Match 47.6%; Score 39; DB 11; Length 249;
 Best Local Similarity 53.8%; Pred. No. 57;
 Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 IGGPPKTRNMSAC 15

Db 148 LDGPHDTRNCEAC 160

RESULT 13

ID Q9D3D6 PRELIMINARY; PRT; 249 AA.
 AC Q9D3D6;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE 5830458K16Rik PROTEIN.
 GN 5830458K16Rik.
 OS Eukaryota (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=THYMUS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Alzava K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamakawa I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirimi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK018021; BAB31041.1; -;
 DR MGD: MGI:1915025; 5830458K16Rik.
 SQ SEQUENCE 249 AA; 28343 MW; 232661EBA18042D4 CRC64;

Query Match 47.6%; Score 39; DB 11; Length 249;
 Best Local Similarity 53.8%; Pred. No. 57;
 Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 IGGPPKTRNMSAC 15

Db 148 LDGPHDTRNCEAC 160

RESULT 14

ID Q9C9Q4 PRELIMINARY; PRT; 651 AA.
 AC Q9C9Q4;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOTHEICAL 71.6 KDA PROTEIN.
 GN F9E10.5.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marziani A.,
 RA Millscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.";
 RL Nature 408:816-820(2000).
 DR EMBL: AC013258; AAG51921.1; -;
 DR InterPro; IPR001623; DnaJ_N.
 DR SMART; SM00271; DnaJ; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 651 AA; 71601 MW; 1FC2139336D27EBC CRC64;

Query Match 47.6%; Score 39; DB 10; Length 651;
 Best Local Similarity 54.5%; Pred. No. 1.5e+02;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 ELIGGPKTRN 11

Db 35 DVFGPPKRRS 45

RESULT 15

ID Q9V9W8 PRELIMINARY; PRT; 815 AA.
 AC Q9V9W8;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CG11518 PROTEIN.
 GN BEST:LD21971 OR CG11518.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003778; AAF57161.1;
DR FlyBase: FBgn0028951; BEST:LD21971.
DR InterPro: IPR001965; PHD.
DR InterPro: IPR002965; P-rich_extensn.
DR InterPro: IPR001412; tRNA-synt_I.
DR Pfam: PF00628; PHD; 1.
DR PRINTS: PR01217; PRICHTEXTENS.
DR SMART: SM00249; PHD; 1.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; UNKNOWN_2.
SQ SEQUENCE 815 AA; 80493 MW; 369FD5A5D34BC136 CRC64;

Query Match 47.6%; Score 39; DB 5; Length 815;
Best Local Similarity 63.6%; Pred. NO. 1.8e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 IGGPPKTRNMS 13
DB 181 MGGPPPMRQMS 191

Search completed: January 31, 2002, 09:36:18
Job time: 787 sec

```

RESULT 1
TSP2_HUMAN
ID TSP2_HUMAN STANDARD; GPC PRT; 1172 AA.
AC P35442;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 2 PRECURSOR.
GN THBS2 OR TSP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=94010892; PubMed=8406456;
RA Labell T.L., Byers P.H.;
RT "Sequence and characterization of the complete human thrombospondin 2
RT cDNA: potential regulatory role for the 3' untranslated region.";
RN Genomics 17:225-229(1993).
[2]
RN SEQUENCE OF 560-1172 FROM N.A.
RX TISSUE=Fibroblast;
RA MEDLINE=92217961; PubMed=1559694;
RA Labell T.L., McGookey Miliewicz D.J., Distèche C.M., Byers P.H.;
RT "Thrombospondin II: partial cDNA sequence, chromosome location, and
RT expression of a second member of the thrombospondin gene family in
RT humans.";
RN Genomics 12:421-429(1992).
RL
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWF DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collabor
CC between the Swiss Institute of Bioinformatics and the EMBL outstat
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for comm
CC entities requires a license agreement (See http://www.isb-sib.ch/anno
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; L12350; AAA03703.1; -
CC EMBL; M81339; -; NOT_ANNOTATED_CDS.
CC PIR; A42173; A42173.
CC HSSP; P00740; 11XA.
CC MM; 188061; -
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR000884; TSP1.

```



```

FT CHAIN 19 1170 THROMBOSPONDIN 2.
FT DOMAIN 19 232 HEPARIN-BINDING (POTENTIAL).
FT FT 375 WFEC.
FT DOMAIN 318 375 TSP TYPE-1 1.
FT DOMAIN 379 430 TSP TYPE-1 2.
FT DOMAIN 435 491 TSP TYPE-1 3.
FT DOMAIN 492 546 TSP TYPE-1 4.
FT DOMAIN 547 587 EGF-LIKE 1.
FT DOMAIN 588 645 EGF-LIKE 2.
FT DOMAIN 646 690 EGF-LIKE 3.
FT DOMAIN 723 758 TSP TYPE-3 1.
FT DOMAIN 759 781 TSP TYPE-3 2.
FT DOMAIN 782 817 TSP TYPE-3 3.
FT DOMAIN 818 840 TSP TYPE-3 4.
FT DOMAIN 841 878 TSP TYPE-3 5.
FT DOMAIN 879 914 TSP TYPE-3 6.
FT DOMAIN 915 950 TSP TYPE-3 7.
FT DOMAIN 951 1170 C-TERMINAL.
FT SITE 926 928 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 266 266 INTERCHAIN (PROBABLE).
FT DISULFID 270 270 INTERCHAIN (PROBABLE).
FT DISULFID 551 562 BY SIMILARITY.
FT DISULFID 556 572 BY SIMILARITY.
FT DISULFID 575 586 BY SIMILARITY.
FT DISULFID 592 608 BY SIMILARITY.
FT DISULFID 599 617 BY SIMILARITY.
FT DISULFID 620 644 BY SIMILARITY.
FT DISULFID 650 663 BY SIMILARITY.
FT DISULFID 657 676 BY SIMILARITY.
FT DISULFID 678 689 BY SIMILARITY.
FT CARBOHYD 151 151 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 455 455 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 582 582 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 708 708 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 936 936 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 535 535 A -> V (IN REF. 3).
FT CONFLICT 748 748 S -> T (IN REF. 3).
SQ SEQUENCE 1170 AA; 129862 MW; 9CFLBFB55B89A051 CRC64;

Query Match 95.1%; Score 78; DB 1; Length 1170;
Best Local Similarity 93.3%; Pred. No. 1.9e-05;
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ELIGGPKTRNWSAC 15
Db 306 ELIGGPKTRNVSAC 320

RESULT 3
TSP2_MOUSE STANDARD; PRT; 1172 AA.
AC Q03350;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 2 PRECURSOR.
GN THB2 OR TSP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92147683; PubMed=1371115;
RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
RA Dixit V.M.;
RT *Characterization of mouse thrombospondin 2 sequence and expression
RT during cell growth and development.*;
RL J. Biol. Chem. 267:3274-3281(1992).
RN [2]

```

```

RP SEQUENCE OF 1-873 FROM N.A.
RX MEDLINE=91302287; PubMed=1712771;
RA Bornstein P., O'Rourke K., Wikstrom K., Wolf F.W., Katz R., Li P.,
RA Dixit V.M.;
RT *A second, expressed thrombospondin gene (Thbs2) exists in the mouse
RT genome.*;
RL J. Biol. Chem. 266:12821-12824(1991).
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 WFEC DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L07803; AAA53064.1; -.
DR EMBL; M64866; AAA40432.1; -.
DR PIR; A42587; A42587.
DR PIR; A39851; A39851.
DR HSP; P00740; IIXA.
DR MGI; 98738; Thbs2.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; WFEC.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 9.
DR Pfam; PF00093; wvc; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00210; TSPN; 3.
DR SMART; SM00214; WVC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50092; TSP1; 3.
DR PROSITE; PS01208; WFEC; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1172 THROMBOSPONDIN 2.
FT DOMAIN 19 232 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 318 375 WFEC.
FT DOMAIN 381 432 TSP TYPE-1 1.
FT DOMAIN 437 493 TSP TYPE-1 2.
FT DOMAIN 494 548 TSP TYPE-1 3.
FT DOMAIN 549 589 EGF-LIKE 1.
FT DOMAIN 590 647 EGF-LIKE 2.
FT DOMAIN 648 692 EGF-LIKE 3.
FT DOMAIN 725 760 TSP TYPE-3 1.
FT DOMAIN 761 783 TSP TYPE-3 2.
FT DOMAIN 784 819 TSP TYPE-3 3.
FT DOMAIN 820 842 TSP TYPE-3 4.
FT DOMAIN 843 880 TSP TYPE-3 5.
FT DOMAIN 881 916 TSP TYPE-3 6.
FT DOMAIN 917 952 TSP TYPE-3 7.
FT DOMAIN 953 980 C-TERMINAL.
FT SITE 928 930 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 266 266 INTERCHAIN (PROBABLE).

```

FT DISULFID 270 270 INTERCHAIN (PROBABLE).
FT DISULFID 553 564 BY SIMILARITY.
FT DISULFID 558 574 BY SIMILARITY.
FT DISULFID 577 588 BY SIMILARITY.
FT DISULFID 594 610 BY SIMILARITY.
FT DISULFID 601 619 BY SIMILARITY.
FT DISULFID 622 646 BY SIMILARITY.
FT DISULFID 652 665 BY SIMILARITY.
FT DISULFID 659 678 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT CARBOHYD 151 151 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 151 151 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 584 584 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1069 1069 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1172 AA; 129911 MW; 7CE8E4E8599822AB CRC64;

Query Match 87.8%; Score 72; DB 1; Length 1172;
Best Local Similarity 93.3%; Pred. No. 0.0002;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ELIGPPKTRNMSAC 15
DB 306 ELIGGPKLTRNMSAC 320
IIIIII IIIIIII

RESULT 4
VE1_HPV41 STANDARD; PRT; 614 AA.
AC P27551;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE REPLICATION PROTEIN E1.
GN E1.
OS Human papillomavirus type 41.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10589;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91253264; PubMed=1645904;
RA Hirt L., Hirsch-Behnam A., de Villiers E.M.;
RT "Nucleotide sequence of human papillomavirus (HPV) type 41: an unusual HPV type without a typical E2 binding site consensus sequence.";
RT Virus Res. 18:179-190(1991).
CC -!- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2 PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH CONTAINS BINDING SITES FOR BOTH PROTEINS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; X56147; CAA39614.1; -
DR PIR; C43550; W1WL41.
DR InterPro; IPR001177; Papillom_E1.
DR Pfam; PF00519; E1; 1.
DR Pfam; PF00524; E1_N; 1.
KW Early protein; DNA replication; Helicase; ATP-binding; DNA-binding; Nucleic acid protein.
FT NP_BIND 443 450 ATP (POTENTIAL).
SQ SEQUENCE 614 AA; 70338 MW; 82EFA0B2A9F01F8 CRC64;

Query Match 53.7%; Score 44; DB 1; Length 614;
Best Local Similarity 53.8%; Pred. No. 5.9;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 LIGGPPKTRNMSA 14
DB 270 MLANPPTRNMPA 282
:::IIIIII

RESULT 5
SSPO_BOVIN STANDARD; PRT; 867 AA.
ID SSPO_BOVIN
AC P98167;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SCO-SPONDIN (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-Ependymocyte;
RX MEDLINE=96338614; PubMed=8743952;
RA Gobron S., Monnerie H., Meinzel R., Creveaux I., Lehmann W., Lamalle D., Dastugue B., Meinzel A.;
RT "SCO-spondin: a new member of the thrombospondin family secreted by the subcommissural organ is a candidate in the modulation of neuronal aggregation.";
RT J. Cell Sci. 109:1053-1061(1996).
RL -!- FUNCTION: INVOLVED IN THE MODULATION OF NEURONAL AGGREGATION.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- TISSUE SPECIFICITY: SUBCOMMISSURAL ORGAN.
CC -!- DEVELOPMENTAL STAGE: EMBRYO.
CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN-FAMILY.
CC -!- SIMILARITY: CONTAINS AT LEAST 4 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS AT LEAST 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS AT LEAST 1 F5/8 TYPE C DOMAIN.
CC -!- SIMILARITY: CONTAINS AT LEAST 3 LDL-RECEPTOR CLASS A DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; X93922; CAA63815.1; -
DR HSP; P01130; IAJJ.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR002919; TIL.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR001007; VWF.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR Pfam; PF00057; ldl_recept_a; 3.
DR Pfam; PF01826; TIL; 1.
DR Pfam; PF00090; tsp_1; 4.
DR Pfam; PF00093; vwc_1; 1.
DR SMART; SM00231; FA58C; 1.
DR SMART; SM00192; LDLA; 3.
DR SMART; SM00209; TSP1; 4.
DR SMART; SM00214; VWC; 1.
DR SMART; SM00011; VWC_def; 1.
DR PROSITE; PS01285; FA58C_1; 1.
DR PROSITE; PS01286; FA58C_2; 1.
DR PROSITE; PS01209; LDLRA_1; 3.
DR PROSITE; PS00068; LDLRA_2; 3.

DR PROSITE; PS50092; TSP1; 4.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain.
 FT NON_TER 1 1
 FT DOMAIN 26 81 TSP TYPE-1 1.
 FT DOMAIN 103 142 EGF-LIKE 1.
 FT DOMAIN 143 180 EGF-LIKE 2.
 FT DOMAIN 185 243 TSP TYPE-1 2.
 FT DOMAIN 344 502 F5/8 TYPE C.
 FT DOMAIN 506 544 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 663 701 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 723 761 LDL-RECEPTOR CLASS A 3.
 FT DOMAIN 762 813 TSP TYPE-1 3.
 FT DOMAIN 814 867 TSP TYPE-1 4.
 FT DISULFID 107 122 BY SIMILARITY.
 FT DISULFID 116 127 BY SIMILARITY.
 FT DISULFID 129 141 BY SIMILARITY.
 FT DISULFID 147 166 BY SIMILARITY.
 FT DISULFID 149 169 BY SIMILARITY.
 FT DISULFID 171 179 BY SIMILARITY.
 FT DISULFID 344 502 BY SIMILARITY.
 FT DISULFID 508 520 BY SIMILARITY.
 FT DISULFID 515 533 BY SIMILARITY.
 FT DISULFID 527 542 BY SIMILARITY.
 FT DISULFID 665 677 BY SIMILARITY.
 FT DISULFID 672 690 BY SIMILARITY.
 FT DISULFID 684 699 BY SIMILARITY.
 FT DISULFID 725 737 BY SIMILARITY.
 FT DISULFID 732 750 BY SIMILARITY.
 FT DISULFID 744 759 BY SIMILARITY.
 FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 867 867
 SQ SEQUENCE 867 AA; 91817 MW; 9538F2108E787B49 CRC64;

Query Match 52.4%; Score 43; DB 1; Length 867;
 Best Local Similarity 57.1%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 LIGGPPKTRNWSAC 15
 I I I I I I I I
 Db 68 LPGGPPSTRETRPC 81

RESULT 6
 PYRP_BACCL STANDARD; PRT; 432 AA.
 AC P41006;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE URACIL PERMEASE (URACIL TRANSPORTER).
 GN PYRP.
 OS Bacillus caldolyticus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 405;
 RX MEDLINE=94266723; PubMed=8206848;
 RA Chim S.Y., Neuhard J.;
 RT "The pyrimidine biosynthesis operon of the thermophile Bacillus
 RT caldolyticus includes genes for uracil phosphoribosyltransferase and
 RT uracil permease.";
 RL J. Bacteriol. 176:3698-3707(1994).
 CC -1- FUNCTION: TRANSPORT OF URACIL IN THE CELL.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE XANTHINE/URACIL PERMEASES FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X76083; CAA53697.1; -.
 DR PIR; S38893; S38893.
 DR InterPro: IPR000444; Xan_ur_permease.
 DR Pfam; PF00860; xan_ur_permease; 1.
 DR PROSITE; PS01116; XANTH_URACIL_PERMASE; 1.
 KW TRANSPORT; Transmembrane.
 FT TRANSMEM 25 45 POTENTIAL.
 FT TRANSMEM 65 85 POTENTIAL.
 FT TRANSMEM 89 109 POTENTIAL.
 FT TRANSMEM 124 144 POTENTIAL.
 FT TRANSMEM 155 175 POTENTIAL.
 FT TRANSMEM 181 201 POTENTIAL.
 FT TRANSMEM 206 226 POTENTIAL.
 FT TRANSMEM 228 248 POTENTIAL.
 FT TRANSMEM 305 325 POTENTIAL.
 FT TRANSMEM 330 350 POTENTIAL.
 FT TRANSMEM 370 390 POTENTIAL.
 FT TRANSMEM 393 413 POTENTIAL.
 SQ SEQUENCE 432 AA; 45182 MW; C015DB9AD515F4F4 CRC64;

Query Match 51.2%; Score 42; DB 1; Length 432;
 Best Local Similarity 87.5%; Pred. No. 9;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 LIGGPPKPT 9
 I I I I I I I I
 Db 284 LLGGPPKPT 291

RESULT 7
 VEL_HPV21 STANDARD; PRT; 603 AA.
 AC P50759;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE REPLICATION PROTEIN E1.
 GN E1.
 OS Human papillomavirus type 21.
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_TaxID=31548;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Delius H.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF
 CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2
 CC PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH
 CC CONTAINS BINDING SITES FOR BOTH PROTEINS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U31779; AAA79396.1; -.
 DR InterPro: IPR001177; Papillom_E1.
 DR Pfam; PF00519; E1; 1.
 DR Pfam; PF00524; E1_N; 1.
 KW Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;
 KW Nuclear protein.

```

FT NP_BIND 431 438 ATP (POTENTIAL).
SQ SEQUENCE 603 AA; 68983 MW; 93006494BEF10339 CRC64;

Query Match
Best Local Similarity 50.0%; Score 41; DB 1; Length 603;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 ELIGPPKTRNSA 14
   ::| ||| ||::|
Db 257 QIISPPKLRNVAA 270

RESULT 8
NQ03_PARDE
ID NQ03_PARDE STANDARD; PRT; 672 AA.
AC P29915;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 3 (EC 1.6.5.3) (NADH
DE DEHYDROGENASE 1, CHAIN 3) (NDH-1, CHAIN 3).
GN NQ03.
OS Paracoccus denitrificans.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Paracoccus.
OX NCBI_TaxID=266;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
RC STRAIN=ATCC 13548;
RX MEDLINE=92296779; PubMed=1605643;
RA Xu X., Matsuno-Yagi A., Yagi T.;
RT "Structural features of the 66-kDa subunit of the energy-transducing
RT NADH-ubiquinone oxidoreductase (NDH-1) of Paracoccus denitrificans."
RL Arch. Biochem. Biophys. 296:40-48(1992).
RN [2]
RP SEQUENCE OF 657-672 FROM N.A.
RX MEDLINE=93136200; PubMed=8422400;
RA Xu X., Matsuno-Yagi A., Yagi T.;
RT "DNA sequencing of the seven remaining structural genes of the gene
RT cluster encoding the energy-transducing NADH-quinone oxidoreductase
RT of Paracoccus denitrificans."
RL Biochemistry 32:968-981(1993).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE - NAD(+) + UBIQUINOL.
CC -1- COPACITOR: MAY BIND TWO 4FE-4S CLUSTER AND ONE 2FE-2S CLUSTER.
CC -1- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 75 KDA SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M84572; AAA25387.1; -.
DR PIR; S23948; S23948.
DR InterPro; IPR000283; Complex1_75k.
DR InterPro; IPR001041; Ferredoxin.
DR InterPro; IPR001467; Molybdopterin.
DR Pfam; PF00111; fer2; 1.
DR Pfam; PF00384; molybdopterin; 1.
DR PROSITE; PS00641; COMPLEX1_75K.1; 1.
DR PROSITE; PS00642; COMPLEX1_75K.2; 1.
DR PROSITE; PS00643; COMPLEX1_75K.3; 1.
KW Oxidoreductase; NAD; Ubiquinone; Iron-sulfur; 4Fe-4S.
FT INIT_MET 0
FT METAL 25 35 -IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 36 36 -IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 47 47 -IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 50 50 -IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 109 109 -IRON-SULFUR (2FE-2S) (POTENTIAL).

FT METAL 112 112 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 118 118 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 157 157 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 160 160 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 163 163 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 207 207 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 672 AA; 73028 MW; EFA5E3A3A70B0560 CRC64;

Query Match
Best Local Similarity 48.8%; Score 40; DB 1; Length 672;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ELIGPPPK 8
   |::|::|::|
Db 53 EVVGPPPK 60

RESULT 9
PIX3_HUMAN
ID PIX3_HUMAN STANDARD; PRT; 302 AA.
AC O75364;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PITUITARY HOMEOBOX 3 (HOMEOBOX PROTEIN PITX3).
DE PITX3 OR PTX3.
GN PITX3 OR PTX3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT ADCC ASN-13.
RC TISSUE=Craniofacial;
RX MEDLINE=98282096; PubMed=9620774;
RA Semina E.V., Ferrell R.E., Mintz-Hittner H.A., Bitoun P.,
RA Alward W.L.M., Reiter R.S., Funkhauser C., Daack-Hirsch S.,
RA Murray J.C.;
RT "A novel homeobox gene PITX3 is mutated in families with autosomal-
RT dominant cataracts and ASDM."
RL Nat. Genet. 19:167-170(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY PLAY A ROLE IN NORMAL ANTERIOR-CHAMBER AND LENS
CC DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN DEVELOPING EYE LENS.
CC -1- DISEASE: MUTATIONS IN PITX3 APPEAR TO BE THE CAUSE OF THE ANTERIOR
CC SEGMENT MESENCHYMAL DYSGENESIS (ASMD) AND AUTOSOMAL-DOMINANT
CC CONGENITAL CATARACTS (ADCC). ASMD INCLUDES ALL MALFORMATIONS
CC INVOLVING THE FIRST (CORNEAL ENDOTHELIUM AND TRABECULAR MESHWORK),
CC SECOND (CORNEAL STROMA) AND THIRD (IRIS STROMA) MESENCHYMAL WAVES
CC OF NEURAL CREST. THE ASMD PHENOTYPE IS CHARACTERIZED BY CORNEAL
CC OPACITIES WITH OR WITHOUT IRIS ADHESIONS IN 100%, CATARACTS OF
CC VARYING SEVERITY IN 100% AND OPTIC-NERVE ABNORMALITIES IN 20% OF
CC AFFECTED INDIVIDUALS. ADCC IS CHARACTERIZED BY DOMINANT
CC TRANSMISSION OF A PHENOTYPE CONSISTING OF BILATERAL CONGENITAL
CC CATARACTS IN A MOTHER AND SON WITHOUT CLINICAL ANTERIOR-SEGMENT
CC ANOMALIES.
CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEOBOX PROTEINS.
CC -1- SIMILARITY: SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

CC EMBL; AF041339; AAC24502.1; -
DR EMBL; BC011642; AAH11642.1; -
DR MIM; 602669; -
DR MIM; 107250; -
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR003654; OAR_domain.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Disease mutation.
FT DNA_BIND 62 121 HOMEBOX.
FT DOMAIN 262 275 OAR.
FT DOMAIN 202 205 POLY-ALA.
FT DOMAIN 240 250 POLY-ALA.
FT DOMAIN 268 272 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT VARIANT 13 13 S -> N (IN ADCC).
FT /FTID=VAR_003767.
SQ SEQUENCE 302 AA; 31832 MW; 1E5259206ABC2E87 CRC64;

Query Match 47.6%; Score 39; DB 1; Length 302;
Best Local Similarity 46.2%; Pred. No. 20;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 IGGPPKTRNMSAC 15
DB 282 VHGPPPAANLSPC 294
: ||| | | |

RESULT 10
PIX3_MOUSE
ID PIX3_MOUSE STANDARD; PRT; 302 AA.
AC O35160;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PITUITARY HOMEBOX 3 (HOMEBOX PROTEIN PITX3).
GN PITX3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo, and Embryonic carcinoma;
RX MEDLINE=97472463; PubMed=9328475;
RA Semina E.V., Reiter R.S., Murray J.C.;
RT "Isolation of a new homeobox gene belonging to the Pitx/Rieg family:
RT expression during lens development and mapping to the aphakia region
RT on mouse chromosome 19.";
RL Hum. Mol. Genet. 6:2109-2116(1997).
RN [2]
RP DEVELOPMENTAL EXPRESSION.
RX MEDLINE=98282096; PubMed=9620774;
RA Semina E.V., Ferrell R.E., Mintz-Hittner H.A., Bitoun P.,
RA Alward W.L.M., Reiter R.S., Funkhauser C., Daack-Hirsch S.,
RA Murray J.C.;
RT "A novel homeobox gene PITX3 is mutated in families with autosomal-
RT dominant cataracts and ASD.";
RL Nat. Genet. 19:167-170(1998).
CC -1- FUNCTION: APPEARS TO BE INVOLVED IN NORMAL EYE ANTERIOR-CHAMBER
CC AND LENS DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN DEVELOPING EYE LENS.
CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED IN THE EYE IN DAY 10
CC POST-CONITUM EMBRYOS. THROUGHOUT EYE DEVELOPMENT, EXPRESSED IN THE
CC LENS PLACODE AND FORMING LENS PIT. FROM DAY 12, ALSO DETECTED IN
CC THE MIDBRAIN REGION, TONGUE, INCISOR PRIMORDIA, CONDENSING
CC MESENCHYME AROUND THE STERNUM AND VERTEBRAE AND IN THE HEAD

CC MUSCLES.
CC -1- DISEASE: MUTATIONS IN PITX3 APPEAR TO BE THE CAUSE OF THE APHAKIA
CC PHENOTYPE, A RECESSIVE HOMOZYGOUS DISEASE CHARACTERIZED BY SMALL
CC EYES AND CLOSED EYELIDS.
CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.
CC "BICOID" SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF005772; AAB87380.1; -
DR HSSP; P06601; IFJL.
DR MGD; MGI:1100498; Pitx3.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR003654; OAR_domain.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DNA_BIND 62 121 HOMEBOX.
FT DOMAIN 262 275 OAR.
FT DOMAIN 202 205 POLY-ALA.
FT DOMAIN 240 250 POLY-ALA.
FT DOMAIN 268 272 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 302 AA; 31714 MW; EB6EF6863B349264 CRC64;

Query Match 47.6%; Score 39; DB 1; Length 302;
Best Local Similarity 46.2%; Pred. No. 20;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 IGGPPKTRNMSAC 15
DB 282 VEGPPPAANLSPC 294
: ||| | | |

RESULT 11
PIX3_RAT
ID PIX3_RAT STANDARD; PRT; 302 AA.
AC P81062;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PITUITARY HOMEBOX 3 (HOMEBOX PROTEIN PITX3).
GN PITX3 OR PTX3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hypothalamus;
RX MEDLINE=98058810; PubMed=9371841;
RA Smidt M.P., van Schaick H.S.A., Lancot C., Tremblay J.T., Cox J.J.,
RA van der Kleij A.A.M., Wolterink G., Drouin J., Burbach J.P.H.;
RT "A homeodomain gene Ptx3 has highly restricted brain expression in
RT mesencephalic dopaminergic neurons.";
RL proc. Natl. Acad. Sci. U.S.A. 94:13305-13310(1997).
RN [2]
RP REVISIONS TO 66; 108; 122 AND 132-135.
RA Smidt M.P.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: APPEARS TO BE INVOLVED IN NORMAL EYE ANTERIOR-CHAMBER
CC AND LENS DEVELOPMENT. TRANSCRIPTION FACTOR WHICH MAY BE INVOLVED
CC IN DEVELOPMENTAL DETERMINATION OF THE MESDA.

```

CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- TISSUE SPECIFICITY: RESTRICTED TO MESENCEPHALIC DOPAMINERGIC
CC SYSTEM (MESDA).
CC -!- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.
CC "BICOID" SUBFAMILY.
CC -!- SIMILARITY: CONTAINS 1 OAR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X74471; CAA52526.1; -.
CC DR EMBL; M96321; AAA47010.1; -.
CC DR PIR; S36493; S36493.
CC DR InterPro; IPR001177; Papillom_E1.
CC DR Pfam; PF00519; E1; 1.
CC DR Pfam; PF00524; E1_N; 1.
CC DR Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;
CC Nuclear protein.
CC KW NP_BIND 432
CC FT NP_BIND 432
CC SQ SEQUENCE 604 AA; 68925 MW; 0A98E84E93B19B9 CRC64;

Query Match 47.6%; Score 39; DB 1; Length 604;
Best Local Similarity 42.9%; Pred. No. 41;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 ELIGPPKTRNMSA 14
Db 258 QILSEPPKLRNVA 271
::: ||| ||::|

RESULT 13
VE1_HPV14
ID VE1_HPV14 STANDARD; PRT; 605 AA.
AC P36721.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE REPLICATION PROTEIN E1.
GN E1.
OS Human papillomavirus type 14.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10605;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94265501; PubMed=8205838;
RA Delli H., Hofmann B.;
RT "Primer-directed sequencing of human papillomavirus types.";
RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).
CC -!- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF
CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2
CC PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH
CC CONTAINS BINDING SITES FOR BOTH PROTEINS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X74467; CAA52502.1; -.
CC DR PIR; S36469; S36469.
CC DR InterPro; IPR001177; Papillom_E1.
CC DR Pfam; PF00519; E1; 1.
CC DR Pfam; PF00524; E1_N; 1.
CC DR Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;
CC Nuclear protein.
CC KW NP_BIND 433
CC FT NP_BIND 433
CC SQ SEQUENCE 605 AA; 69302 MW; 6702915505AA6491 CRC64;

Query Match 47.6%; Score 39; DB 1; Length 302;
Best Local Similarity 46.2%; Pred. No. 20;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 IGGPPKTRNMSAC 15
Db 282 VPGPPPAANLSPC 294
::: ||| ||::|

RESULT 12
VE1_HPV25
ID VE1_HPV25 STANDARD; PRT; 604 AA.
AC Q02049.
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE REPLICATION PROTEIN E1.
GN E1.
OS Human papillomavirus type 25.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10609;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94265501; PubMed=8205838;
RA Delli H., Hofmann B.;
RT "Primer-directed sequencing of human papillomavirus types.";
RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).
CC -!- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF
CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2
CC PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH
CC CONTAINS BINDING SITES FOR BOTH PROTEINS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ011005; CNA09455.2; -.
CC DR HSP; P06601; IFJL.
CC DR InterPro; IPR001356; Homeobox.
CC DR InterPro; IPR003654; OAR_domain.
CC DR Pfam; PF00046; homeobox; 1.
CC DR PRINTS; PR00024; HOMEBOX.
CC DR SMART; SM00389; HOX; 1.
CC DR PROSITE; PS00027; HOMEBOX_1; FALSE_NEG.
CC DR PROSITE; PS00071; HOMEBOX_2; 1.
CC KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
CC FT DNA_BIND 62 121
CC FT DOMAIN 262 275
CC FT DOMAIN 202 205
CC FT DOMAIN 240 250
CC FT DOMAIN 268 272
CC FT DOMAIN 302 AA; 31728 MW; DE0E81863C44E714 CRC64;

```

Query Match 47.6%; Score 39; DB 1; Length 605;
Best Local Similarity 42.9%; Pred. No. 41;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 ELIGGPPKTRNMSA 14
::: ||| ||::|

Db 259 QILSEPPKLRNVA 272

RESULT 14

VE1_HPV20 STANDARD; PRT; 605 AA.

AC P50758;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE REPLICATION PROTEIN E1.

GN E1.

OS Human papillomavirus type 20.

OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;

OC Papillomavirus.

OX NCBI_TaxID=31547;

RN [1]

RP SEQUENCE FROM N.A.

RA Delius H.;

RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF

VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2

PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH

CONTAINS BINDING SITES FOR BOTH PROTEINS.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U31778; AAA79389.1; -.

DR InterPro; IPR001177; Papillom_E1.

DR Pfam; PF00519; E1; 1.

DR Pfam; PF00524; EL_N; 1.

KW Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;

Nuclear protein.

FT NP_BIND 433 440 ATP (POTENTIAL).

SQ SEQUENCE 605 AA; 69036 MW; ACD36510DC4BE958 CRC64;

Query Match 47.6%; Score 39; DB 1; Length 605;
Best Local Similarity 42.9%; Pred. No. 41;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 ELIGGPPKTRNMSA 14
::: ||| ||::|

Db 259 QILSEPPKLRNVA 272

RESULT 15

YPI7_CAEL

ID YPI7_CAEL STANDARD; PRT; 731 AA.

AC Q18262;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE HYPOTHETICAL 83.8 KDA PROTEIN C27F2.7 IN CHROMOSOME III.

GN C27F2.7.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN SEQUENCE FROM N.A.
RP STRAIN=BRISTOL N2;
RA Kirsten J. ;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO S.FOMBE SPAC1B3.08.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U40419; AAA81427.1; -.

DR WormPep; C27F2.7; CE01174.

DR InterPro; IPR001810; F-box.

DR Pfam; PF00646; F-box; 1.

DR SMART; SM00256; FBOX; 1.

DR PROSITE; PS50181; FBOX; 1.

KW Hypothetical protein.

SQ SEQUENCE 731 AA; 83824 MW; 723226803BF4757D CRC64;

Query Match 47.6%; Score 39; DB 1; Length 731;
Best Local Similarity 64.3%; Pred. No. 49;
Matches 9; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

QY 4 GGPP--KTRNMSAC 15

||| |||||

Db 389 GKPPPVKTRNTSRC 402

Search completed: January 31, 2002, 09:36:51
Job time: 770 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:23:56 ; Search time 43.81 seconds
(without alignments)
7.705 Million cell updates/sec

Title: US-09-536-087-9

Perfect score: 82

Sequence: 1 ELIGGPPKTRNMSAC 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTRUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	1172	1	US-08-313-288B-19
2	42	51.2	432	2	US-08-677-049-8
3	39	47.6	302	4	US-08-957-351-3
4	39	47.6	302	4	US-08-957-351-7
5	39	47.6	302	4	US-08-957-351-30
6	38	46.3	434	2	US-08-677-049-7
7	37	45.1	439	1	US-08-333-358-10
8	37	45.1	439	1	US-08-463-694-10
9	37	45.1	439	1	US-08-694-501-10
10	37	45.1	440	2	US-08-484-200-4
11	36.5	44.5	72	1	US-08-439-818A-21
12	36.5	44.5	72	2	US-08-751-965-21
13	36.5	44.5	72	2	US-08-738-975-21
14	36.5	44.5	72	2	US-08-728-626-21
15	36.5	44.5	72	3	US-08-808-599A-21
16	36.5	44.5	674	1	US-08-317-522A-3
17	36.5	44.5	674	1	US-08-439-818A-3
18	36.5	44.5	674	2	US-08-751-965-3
19	36.5	44.5	674	2	US-08-738-975-3
20	36.5	44.5	674	2	US-08-728-626-3
21	36.5	44.5	674	3	US-08-808-599A-3
22	36.5	44.5	749	1	US-08-317-522A-2
23	36.5	44.5	749	1	US-08-439-818A-2
24	36.5	44.5	749	2	US-08-738-975-2
25	36.5	44.5	749	2	US-08-751-965-2
26	36.5	44.5	749	2	US-08-728-626-2
27	36.5	44.5	749	3	US-08-808-599A-2

Sequence 12, Appl
Sequence 72, Appl
Sequence 2, Appl
Sequence 6, Appl
Sequence 2, Appl
Sequence 23, Appl
Sequence 29, Appl
Sequence 25, Appl
Sequence 19, Appl
Sequence 21, Appl
Sequence 27, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 3, Appl
Sequence 80, Appl
Sequence 81, Appl
Sequence 83, Appl

ALIGNMENTS

RESULT 1
US-08-313-288B-19 74 4 US-09-267-177-12
; Sequence 19, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; NUMBER OF INVENTIONS: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1172 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-313-288B-19

Query Match 100.0%; Score 82; DB 1; Length 1172;
Best Local Similarity 100.0%; Pred. No. 4e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ELIGGPPKTRNMSAC 15
|
DB 306 ELIGGPPKTRNMSAC 320

```

RESULT 2
US-08-677-049-8
; Sequence 8, Application US/08677049
; Patent No. 5858707
; GENERAL INFORMATION:
; APPLICANT: Guimaraes, M. Jorge
; APPLICANT: Bazan, J. Fernando
; APPLICANT: McClanahan, Terrill K.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,049
; FILING DATE: 03-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/000,788
; FILING DATE: 03-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Region
; LOCATION: 115..144
; OTHER INFORMATION: /note= "Encompasses TM 4 of Figure
; OTHER INFORMATION: 4"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 325..359
; OTHER INFORMATION: /note= "Encompasses TM 9 of Figure
; OTHER INFORMATION: 4"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 363..390
; OTHER INFORMATION: /note= "Encompasses TM 10 of Figure
; OTHER INFORMATION: 4"
US-08-677-049-8
Query Match 51.2%; Score 42; DB 2; Length 432;
Best Local Similarity 87.5%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 2 LIGGPPKT 9
Db 284 LIGGPPKT 291
I:|||||

RESULT 3
US-08-957-351-3
; Sequence 3, Application US/08957351
; Patent No. 6306586
; GENERAL INFORMATION:
; APPLICANT: Semina, Elena
; APPLICANT: Murray, Jeffrey C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,351
; FILING DATE: 24-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-024.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-957-351-3
Query Match 47.6%; Score 39; DB 4; Length 302;
Best Local Similarity 46.2%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Oy 3 IGGPPKTRNMSAC 15
Db 282 VPGPPPAANLSPC 294
: ||| | | |

RESULT 4
US-08-957-351-7
; Sequence 7, Application US/08957351
; Patent No. 6306586
; GENERAL INFORMATION:
; APPLICANT: Semina, Elena
; APPLICANT: Murray, Jeffrey C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
US-08-957-351-7
Query Match 47.6%; Score 39; DB 4; Length 302;
Best Local Similarity 46.2%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Oy 3 IGGPPKTRNMSAC 15
Db 282 VPGPPPAANLSPC 294
: ||| | | |
```


;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/957,351
;; FILING DATE: 24-OCT-1997
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Arnold, Beth E.
;; REGISTRATION NUMBER: 35,430
;; REFERENCE/DOCKET NUMBER: UIA-024.01
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617-832-1000
;; TELEFAX: 617-832-7000
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 302 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-957-351-7

Query Match 47.6%; Score 39; DB 4; Length 302;
Best Local Similarity 46.2%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 IGGPKTRNMSAC 15
: ||| |
Db 282 VHGPAAANLSPC 294

RESULT 5
US-08-957-351-30
;; Sequence 30, Application US/08957351
;; Patent No. 6306586
;; GENERAL INFORMATION:
;; APPLICANT: Semina, Elena
;; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
;; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS
;; NUMBER OF SEQUENCES: 33
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
;; STREET: One Post Office Square
;; CITY: Boston
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02109-2170
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/957,351
;; FILING DATE: 24-OCT-1997
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Arnold, Beth E.
;; REGISTRATION NUMBER: 35,430
;; REFERENCE/DOCKET NUMBER: UIA-024.01
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617-832-1000
;; TELEFAX: 617-832-7000
;; INFORMATION FOR SEQ ID NO: 30:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 302 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-957-351-30

Query Match 47.6%; Score 39; DB 4; Length 302;

Best Local Similarity 46.2%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
Qy 3 IGGPKTRNMSAC 15
: ||| |
Db 282 VHGPAAANLSPC 294
RESULT 6
US-08-677-049-7
;; Sequence 7, Application US/08677049
;; Patent No. 5858707
;; GENERAL INFORMATION:
;; APPLICANT: Guimaraes, M. Jorge
;; APPLICANT: Bazan, J. Fernando
;; APPLICANT: McClanahan, Terrill K.
;; APPLICANT: Zlotnik, Albert
;; TITLE OF INVENTION: PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
;; TITLE OF INVENTION: NUCLEIC ACIDS; ANTIBODIES
;; NUMBER OF SEQUENCES: 12
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: DNAX Research Institute
;; STREET: 901 California Avenue
;; CITY: Palo Alto
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94304-1104
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/677,049
;; FILING DATE: 03-JUL-1996
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 60/000,788
;; FILING DATE: 03-JUL-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Ching, Edwin P.
;; REGISTRATION NUMBER: 34,090
;; REFERENCE/DOCKET NUMBER: DX0511
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-852-9196
;; TELEFAX: 415-496-1200
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 434 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; FEATURE:
;; NAME/KEY: Region
;; LOCATION: 117..146
;; OTHER INFORMATION: /note= "Encompasses TM 4 of Figure
;; OTHER INFORMATION: 4"
;; FEATURE:
;; NAME/KEY: Region
;; LOCATION: 330..364
;; OTHER INFORMATION: /note= "Encompasses TM 9 of Figure
;; OTHER INFORMATION: 4"
;; FEATURE:
;; NAME/KEY: Region
;; LOCATION: 368..395
;; OTHER INFORMATION: /note= "Encompasses TM 10 of Figure
;; OTHER INFORMATION: 4"
US-08-677-049-7

Query Match 46.3%; Score 38; DB 2; Length 434;
Best Local Similarity 87.5%; Pred. No. 60;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LIGGPPKT 9

|||||

Db 289 LIGGPPTT 296

RESULT 7

US-08-333-358-10

; Sequence 10, Application US/08333358

; Patent No. 5571696

; GENERAL INFORMATION:

; APPLICANT: EVANS Ph.D., RONALD M.

; APPLICANT: MANGELSDORF Ph.D., DAVID J.

; APPLICANT: ONG Ms., ESTELITA S.

; APPLICANT: ORO Ph.D., ANTHONY E.

; APPLICANT: BORGMEYER Ph.D., UWE K.

; APPLICANT: GIGUERE Ph.D., VINCENT NMN

; APPLICANT: YAO Mr., TSO-PANG NMN

; TITLE OF INVENTION: NOVEL RECEPTORS

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark

; STREET: 444 So. Flower St., Suite 2000

; CITY: Los Angeles

; STATE: CA

; COUNTRY: US

; ZIP: 90071-2921

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/333,358

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/761,068

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Reiter Ph.D., Stephen E.

; REGISTRATION NUMBER: 31192

; REFERENCE/DOCKET NUMBER: P31 8936

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 535-9001

; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 439 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-333-358-10

Query Match

Best Local Similarity 45.1%; Score 37; DB 1; Length 439;

Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 ELIGGPPKTRNMSAC 15

|||||

Db 27 ELNGGPEHTLPSSC 41

RESULT 8

US-08-463-694-10

; Sequence 10, Application US/08463694

; Patent No. 5696233

; GENERAL INFORMATION:

; APPLICANT: EVANS Ph.D., RONALD M.

; APPLICANT: MANGELSDORF Ph.D., DAVID J.

; APPLICANT: ONG Ms., ESTELITA S.

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/694,501
; FILING DATE: 07-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/333,358
; FILING DATE:
; APPLICATION NUMBER: US/07/761,068
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter Ph.D., Stephen E.
; REGISTRATION NUMBER: 31192
; REFERENCE/DOCKET NUMBER: P31 8936
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 439 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-694-501-10

```

```

Query Match 45.18; Score 37; DB 1; Length 439;
Best Local Similarity 53.3%; Pred. No. 90;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

```

```

Qy 1 ELIGPPKTRNMSAC 15
| | | | | | | |
Db 27 ELGGPHTLPSSC 41

```

```

RESULT 10
US-08-484-200-4
; Sequence 4, Application US/08484200
; Patent No. 5861274
; GENERAL INFORMATION:
; APPLICANT: EVANS, RONALD M.
; APPLICANT: FORMAN, BARRY M.
; APPLICANT: KLEWER, STEVEN A.
; APPLICANT: ONG, ESTELITA S.
; TITLE OF INVENTION: NOVEL MEMBERS OF THE STEROID/THYROID
; TITLE OF INVENTION: SUPERFAMILY AND USES THEREFOR
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
; STREET: 444 SOUTH FLOWER STREET, SUITE 2000
; CITY: LOS ANGELES
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,200
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: REITER, STEPHEN E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9971
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-1995
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 4:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 440 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-484-200-4

```

```

Query Match 45.18; Score 37; DB 2; Length 440;
Best Local Similarity 53.3%; Pred. No. 91;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

```

```

Qy 1 ELIGPPKTRNMSAC 15
| | | | | | | |
Db 27 ELGGPHTLPSSC 41

```

```

RESULT 11
US-08-439-818A-21
; Sequence 21, Application US/08439818A
; Patent No. 5654145
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting
; TITLE OF INVENTION: Proteins
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,818A
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/317,522
; FILING DATE: 04-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 1563
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-439-818A-21

```

```

Query Match 44.5%; Score 36.5; DB 1; Length 72;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

```

```

Qy 4 GGPPKTRNMSAC 15
| | | | | | | |
Db 48 GGPPST---SAC 56

```

```

RESULT 12
US-08-751-965-21
; Sequence 21, Application US/08751965
; Patent No. 5858360

```

GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting
; TITLE OF INVENTION: Proteins
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,965
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/439,818
; FILING DATE: 12-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 2252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-751-965-21

Query Match 44.5%; Score 36.5; DB 2; Length 72;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

Qy 4 GPPKTRNSAC 15
Db 48 GPPST---SAC 56
RESULT 13
US-08-738-975-21
; Sequence 21, Application US/08738975
; Patent No. 5880267
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting
; TITLE OF INVENTION: Proteins
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,975
; FILING DATE: herewith
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/439,818
; FILING DATE: 05-Dec-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 2251
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-738-975-21

Query Match 44.5%; Score 36.5; DB 2; Length 72;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

Qy 4 GPPKTRNSAC 15
Db 48 GPPST---SAC 56
RESULT 14
US-08-728-626-21
; Sequence 21, Application US/08728626
; Patent No. 5910451
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting
; TITLE OF INVENTION: Proteins
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,626
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/439,818
; FILING DATE: 12-MAY-1995
; APPLICATION NUMBER: US 08/317,522
; FILING DATE: 04-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 1563
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-728-626-21

Query Match 44.5%; Score 36.5; DB 2; Length 72;

Best Local Similarity 66.7%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY 4 GGPKTRNMSAC 15
Db 48 GGPEST---SAC 56

RESULT 15
US-08-808-599A-21
; Sequence 21, Application US/08080599A
; Patent No. 6111089
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin, Trophinin-Assisting
; TITLE OF INVENTION: Proteins and Methods to Inhibit Implantation
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/808,599A
; FILING DATE: 28-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/317,522
; FILING DATE: 04-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/439,818
; FILING DATE: 12-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 2256
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-808-599A-21

Query Match 44.5%; Score 36.5; DB 3; Length 72;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY 4 GGPKTRNMSAC 15
Db 48 GGPEST---SAC 56

Search completed: January 31, 2002, 09:23:57
Job time: 222 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:23:07 ; Search time 96.28 seconds
(without alignments)
4.616 Million cell updates/sec

Title: US-09-536-087-11
Perfect score: 32
Sequence: 1 CSVTVG 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	96.9	208	22 AAB47001	Plant D-like cyclin
2	31	96.9	212	21 AAY44337	Arabidopsis thaliana
3	29	90.6	405	21 AAB52063	Human secreted pro
4	29	90.6	4302	17 AAU00870	Polycystic kidney
5	29	90.6	4302	19 AAU00870	Human PKD1 polypep
6	29	90.6	4302	19 AAU00870	Human PKD1 protein
7	29	90.6	4302	19 AAU00870	Human polycystin
8	29	90.6	4302	17 AAY92719	Polycystic kidney
9	29	90.6	4303	21 AAY95558	Human polycystin-1
10	29	90.6	4339	16 AAR75916	Polycystic kidney
11	29	90.6	4339	17 AAR87539	Polycystic kidney

12	28	87.5	29	22	AAM18776	Peptide #5210 enco
13	28	87.5	29	22	AAW31282	Peptide #5319 enco
14	28	87.5	282	20	AAY36917	Amino acid sequenc
15	28	87.5	305	22	AAG85037	Shrimp white spot
16	28	87.5	343	20	AAY56771	C. trachomatis ser
17	28	87.5	391	20	AAV35319	Chlamydia pneumoni
18	28	87.5	479	22	AAG81189	Mycobacterium tube
19	28	87.5	485	22	AAB95017	Human protein sequ
20	28	87.5	2161	20	AAW87724	Rupestria stem pit
21	28	87.5	3135	15	AAW57474	P. falciparum tran
22	28	87.5	3135	21	AAB18223	Plasmodium falcipa
23	27	84.4	6	13	AAR21723	Peptide promoting
24	27	84.4	6	13	AAR25308	Hexapeptide derive
25	27	84.4	6	14	AAR32434	Human thrombospond
26	27	84.4	6	15	AAR43997	Thrombospondin reg
27	27	84.4	6	17	AAW17307	Anti-haemorrhagic
28	27	84.4	6	17	AAW05010	Thrombospondin-der
29	27	84.4	6	18	AAW22653	Thrombospondin ana
30	27	84.4	6	18	AAW01540	Anti-inflammatory
31	27	84.4	6	20	AAV49510	Human METH1 and ME
32	27	84.4	6	20	AAV21701	EP-927558 peptide
33	27	84.4	6	20	AAV17779	Human thrombospond
34	27	84.4	6	20	AAW67723	Radioimaging pepti
35	27	84.4	6	21	AAB21269	ADAMTS internal TS
36	27	84.4	6	21	AAW00046	Active region of h
37	27	84.4	6	21	AAW02641	Human thrombospond
38	27	84.4	6	22	AAE04974	Thrombospondin (TS
39	27	84.4	6	22	AAE05004	Human prote-oncoge
40	27	84.4	6	22	AAE00916	Conserved heparin-
41	27	84.4	6	22	AAW60582	Thrombospondin (TS
42	27	84.4	6	22	AAW60600	Trap/properdin/cir
43	27	84.4	6	22	AAV72567	Adhesive domain of
44	27	84.4	6	22	AAV72570	Type 1 repeat doma
45	27	84.4	6	22	AAV72584	D-Orientation_pept

ALIGNMENTS

RESULT 1
AAB47001
ID AAB47001 standard; Protein; 208 AA.
XX
AC AAB47001;
DT 22-MAR-2001 (first entry)
XX
DE Plant D-like cyclin inhibitor BRO4.
XX
KW Plant; D-like cyclin inhibitor gene; BRO4; hyperplastic; variant;
KW growth rate; dividing cells; inactivation; protoplast; seed;
KW root cell; meristem; leaf.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT Protein 13..208
FT /label= BRO4
XX
PN WO200069883-A1.
XX
PD 23-NOV-2000.
XX
PF 15-MAY-2000; 2000WO-US13379.
XX
PR 14-MAY-1999; 99US-0134373.
XX
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
PI Roberts J, Kelly B;
XX
DR WPI; 2001-024998/03.
DR N-PSDB; AAC85201.

XX Functionally inactivating expression of plant D-like cyclin inhibitor
PT gene for producing a hyperplastic variant plant, modulating the growth
PT and/or yield of plants, and increasing the proportion of dividing cells
XX
XX Claim 15; Page 41-42; 50pp; English.
XX
CC The sequence given in AAB47001 represents a plant D-like cyclin
CC inhibitor, BR04. The DNA encoding this sequence may be used to
CC produce a hyperplastic variant plant, increase the growth rate of
CC a plant, or increase the proportion of dividing cells in a plant cell
CC population, relative to a wild-type plant, by functionally
CC inactivating the expression of a plant D-like cyclin inhibitor
CC gene in a plant. The DNA sequence is homologous to a sequence present
CC in a D-like cyclin inhibitor gene and when integrated at the
CC corresponding locus, functionally inactivates plant D-like cyclin
CC inhibitor protein expression. BR04 is useful for producing
CC hyperplastic variant plants, increasing the growth rate of a plant
CC and for increasing the proportion of dividing cells in a plant cell
CC population comprising protoplast, seeds, root cells, meristem cells
CC or leaf cells.
XX
XX Sequence 208 AA;
SQ

Query Match 96.9%; Score 31; DB 22; Length 208;
Best Local Similarity 83.3%; Pred. No. 67;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
||||:|
Db 71 csvtig 76

RESULT 2
AAY44337
ID AAY44337 standard; Protein; 212 AA.
XX
AC AAY44337;
XX
XX 29-FEB-2000 (first entry)
XX
XX Arabidopsis thaliana CDK inhibitor, ICN2.
XX
XX Cyclin-Dependent Kinase inhibitor; CDK; Interactor of Cyclin 2;
KW ICN2; Cdc2a; D-class cyclin; CycD1; CycD2; CycD3; morphogenesis;
KW antisense construct; tissue-specific promoter; transgenic plant;
KW male sterility.
XX
XX Unidentified.
XX
XX WO9964599-A1.
XX
XX 16-DEC-1999.
XX
XX 08-JUN-1999; 99WO-CA00532.
XX
XX 08-JUN-1998; 98CA-2235978.
XX
XX 31-DEC-1998; 98CA-2256121.
XX
XX (MTAC) AGRIC & AGRIFOOD CANADA.
XX (UYSA-) UNIV SASKATCHEWAN TECHNOLOGIES INC.
XX (CANA) NAT RES COUNCIL CANADA.
XX
XX Wang H, Fowke LC, Crosby WL;
XX
XX WPI; 2000-097540/08.
XX
XX N-PSDB; AAZ229417.
XX
XX Modifying plant cell development using nucleic acid encoding inhibitor
PT of cyclin-dependent kinase, or corresponding antisense sequence, e.g.
PT for inducing male sterility -
XX

PS Disclosure; Fig 7; 58pp; English.
XX
CC The present protein sequence is ICN2, which inhibits A. thaliana
CC Cyclin-Dependent Kinase (CDK). Interactor of Cyclin 2 (ICN2)
CC interacts with Cdc2a, D-class cyclins, CycD1, CycD2 and CycD3 and
CC shares functional and sequence similarity with IKK1. Growth,
CC morphogenesis, multiplication, enlargement, differentiation and
CC maturation of plant cells can be modified by transforming them with
CC nucleic acid encoding CDK inhibitor or antisense construct complementary
CC to the inhibitor gene, operably linked to a tissue-specific promoter.
CC The transgenic plants exhibit alteration of traits such as petals, male
CC sterility and ability to set seeds.
XX
XX Sequence 212 AA;
SQ

Query Match 96.9%; Score 31; DB 21; Length 212;
Best Local Similarity 83.3%; Pred. No. 58;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
||||:|
Db 75 csvtig 80

RESULT 3
AAB52063
ID AAB52063 standard; Protein; 405 AA.
XX
AC AAB52063;
XX
XX 21-FEB-2001 (first entry)
XX
XX Human secreted protein sequence encoded by gene 2 SEQ ID NO:112.
XX
XX Human; secreted protein; cytostatic; immunosuppressive; nontropic;
KW neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antiulcer; vulnerary; anticonvulsant; antibacterial;
KW antifungal; antiparasitic; cardiant; cancer; immune disease; allergy;
KW cardiovascular disorder; wound healing; infection; neurological disease.
XX
XX Homo sapiens.
XX
XX WO200061596-A1.
XX
XX 19-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US08983.
XX
XX 09-APR-1999; 99US-0128703.
XX
XX 20-JAN-2000; 2000US-0176068.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Ruben SM, Komatsoulis G;
XX
XX WPI; 2000-611865/58.
XX
XX N-PSDB; AAC95522.
XX
XX Fifty nucleic acid molecules encoding human secreted proteins, useful
PT in the prevention, treatment and diagnosis of cancer, immune disorders,
PT cardiovascular disorders and neurological diseases -
XX
XX Disclosure; Page 11; 505pp; English.
XX
XX Polynucleotide sequences AAC95521 - AAC95570 represent cDNA encoding 50
CC human secreted proteins AAB52012 - AAB52161. Sequences AAB52062 -
CC AAB52103 represent alternative polypeptides encoded by the genes, and
CC amino acid sequences with which they share homology. The genes and
CC proteins have activities dependent on the tissues and cells in which they
CC are expressed. Examples of their activities include cytostatic;
CC immunosuppressive; nontropic; neuroprotective; antiviral; antiallergic;

CC hepatotropic; antidiabetic; antiinflammatory; antiulcer; vulneryary;
 CC anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiant.
 CC The secreted proteins, polypeptides, antagonists and agonists may be
 CC useful in treating, preventing and/or diagnosing diseases and disorders
 CC such as cancer, particularly breast and ovarian cancer, and other cancers
 CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,
 CC liver, lung, or urogenital. Immune disorders such as Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemias; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections may also be treated using the
 CC proteins and polynucleotides of the invention. Sequences AAC95512 -
 CC AAC95520 and AAB52011 are used in the isolation and characterisation of
 CC the proteins and polynucleotides of the invention.

XX Sequence 405 AA;

Query Match 90.68; Score 29; DB 21; Length 405;
 Best Local Similarity 83.38; Pred. No. 3.1e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CSVTVG 6
 |:||||
 Db 141 cavtvq 146

RESULT 4

AAW00870
 ID AAW00870 standard; Protein; 4302 AA.

XX AC AAW00870;

XX 02-FEB-1997 (first entry)

DE Polycystic kidney disease 1 (PKD1) polypeptide.

KW Adult polycystic kidney disease; APKD; PKD1 gene; diagnosis;
 KW therapy; polycystin.

XX Homo sapiens.

Key	Location/Qualifiers
Peptide	1..23
Protein	/label= Sig_peptide
Region	24..4303
	/label= Mat_protein
	72..125
	/label= LRR
	/note= "leucine-rich repeat region"
Domain	2580..2600
	/label= TM1
Domain	/note= "transmembrane domain 1"
	2693..2713
	/label= TM2
Domain	/note= "transmembrane domain 2"
	3075..3095
	/label= TM3
Domain	/note= "transmembrane domain 3"
	3281..3301
	/label= TM4
Domain	/note= "transmembrane domain 4"
	3323..3343
	/label= TM5
Domain	/note= "transmembrane domain 5"
	3559..3579
	/label= TM6
Domain	/note= "transmembrane domain 6"
	3582..3612
	/label= TM7
	/note= "transmembrane domain 7"

Domain	3669..3689
	/label= TM8
Misc-difference 50	/note= "transmembrane domain 8"
Misc-difference 89	/label= N-glycosylation_site
Misc-difference 107	/label= N-glycosylation_site
Misc-difference 112	/label= N-glycosylation_site
Misc-difference 187	/label= N-glycosylation_site
Misc-difference 621	/label= N-glycosylation_site
Misc-difference 632	/label= N-glycosylation_site
Misc-difference 746	/label= N-glycosylation_site
Misc-difference 810	/label= N-glycosylation_site
Misc-difference 841	/label= N-glycosylation_site
Misc-difference 854	/label= N-glycosylation_site
Misc-difference 890	/label= N-glycosylation_site
Misc-difference 921	/label= N-glycosylation_site
Misc-difference 1004	/label= N-glycosylation_site
Misc-difference 1034	/label= N-glycosylation_site
Misc-difference 1072	/label= N-glycosylation_site
Misc-difference 1113	/label= N-glycosylation_site
Misc-difference 1178	/label= N-glycosylation_site
Misc-difference 1194	/label= N-glycosylation_site
Misc-difference 1240	/label= N-glycosylation_site
Misc-difference 1269	/label= N-glycosylation_site
Misc-difference 1336	/label= N-glycosylation_site
Misc-difference 1348	/label= N-glycosylation_site
Misc-difference 1382	/label= N-glycosylation_site
Misc-difference 1450	/label= N-glycosylation_site
Misc-difference 1455	/label= N-glycosylation_site
Misc-difference 1474	/label= N-glycosylation_site
Misc-difference 1518	/label= N-glycosylation_site
Misc-difference 1541	/label= N-glycosylation_site
Misc-difference 1554	/label= N-glycosylation_site
Misc-difference 1563	/label= N-glycosylation_site
Misc-difference 1647	/label= N-glycosylation_site
Misc-difference 1661	/label= N-glycosylation_site
Misc-difference 1733	/label= N-glycosylation_site
Misc-difference 1791	/label= N-glycosylation_site
	/label= N-glycosylation_site


```
Query Match      90.6%; Score 29; DB 19; Length 4302;
Best Local Similarity 83.3%; Pred. No. 2.7e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 1270 ctvtvg 1275

RESULT 6
AAW23830
ID AAW23830 standard; Protein; 4302 AA.
XX AC AAW23830;
XX DT 08-MAY-1998 (first entry)
XX DE Human PKD1 protein.
XX DE PKD1; polycystin; polycystic renal degeneration; diagnosis; epitope;
KW therapy.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FH Region 26..270
XX FH /note= "contains epitope recognised by a
XX FH /note= "PKD1-specific antibody"
XX FH Region 26..480
XX FH /note= "contains epitope recognised by a
XX FH /note= "PKD1-specific antibody"
XX FH Region 361..540
XX FH /note= "contains epitope recognised by a
XX FH /note= "PKD1-specific antibody"
XX FH Region 480..700
XX FH /note= "contains epitope recognised by a
XX FH /note= "PKD1-specific antibody"
XX FH Region 541..840
XX FH /note= "contains epitope recognised by a
XX FH /note= "PKD1-specific antibody"
XX FH Region 700..1100
XX FH /note= "contains epitope recognised by a
XX FH /note= "PKD1-specific antibody"
XX FH Region 1011..1220
XX FH /note= "contains epitope recognised by a
XX FH /note= "PKD1-specific antibody"
XX FH Region 2161..2370
XX FH /note= "contains epitope recognised by a
XX FH /note= "PKD1-specific antibody"
XX FH Region 2723..2931
XX FH /note= "contains epitope recognised by a
XX FH /note= "PKD1-specific antibody"
XX FH Region 2850..3000
XX FH /note= "contains epitope recognised by a
XX FH /note= "PKD1-specific antibody"
XX FH Region 2932..3067
XX FH /note= "contains epitope recognised by a
XX FH /note= "PKD1-specific antibody"
XX FH Region 3100..3280
XX FH /note= "contains epitope recognised by a
XX FH /note= "PKD1-specific antibody"
XX FH Region 3200..3400
XX FH /note= "contains epitope recognised by a
XX FH /note= "PKD1-specific antibody"
XX FH Region 3311..3603
XX FH /note= "contains epitope recognised by a
XX FH /note= "PKD1-specific antibody"
XX FH Region 4090..4302
XX FH /note= "contains epitope recognised by a
XX FH /note= "PKD1-specific antibody"
XX PN DE19650758-C1.

Query Match      90.6%; Score 29; DB 19; Length 4302;
Best Local Similarity 83.3%; Pred. No. 2.7e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 1270 ctvtvg 1275

RESULT 7
AAW92719
ID AAW92719 standard; Protein; 4302 AA.
XX AC AAW92719;
XX DT 29-AUG-2000 (first entry)
XX DE Human polycystin.
XX KW Polycystic kidney disease 1; PKD1; polycystin; chromosome 16p13.3;
XX KW epitope; transmembrane; antibody; cell adhesion.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FH Peptide 27..360
XX FH /label= epitope
XX FH /note= "loop region of transmembrane domain"
XX FH Domain 843..1200
XX FH /label= Ig-like
XX FH Domain 1205..1625
XX FH /label= Ig-like
XX FH Domain 1626..2136
XX FH /label= Ig-like
XX FH Peptide 2166..2599
XX FH /label= epitope
XX FH /note= "loop region of transmembrane domain"
XX FH Peptide 2621..2710
XX FH /label= epitope
XX FH /note= "loop region of transmembrane domain"
XX FH Peptide 2734..3094
XX FH /label= epitope
XX FH /note= "loop region of transmembrane domain"
XX FH Peptide 3166..3300
XX FH /label= epitope
XX FH /note= "loop region of transmembrane domain"
```

FT	Peptide	3364..3578	
FT		/label= epitope	
FT		/note= "loop region of transmembrane domain"	
FT	Peptide	3623..3688	
FT		/label= epitope	
FT		/note= "loop region of transmembrane domain"	
FT	Peptide	3710..3914	
FT		/label= epitope	
FT		/note= "loop region of transmembrane domain"	
FT	Peptide	3931..4046	
FT		/label= epitope	
FT		/note= "loop region of transmembrane domain"	
FT	Peptide	4097..4302	
FT		/label= epitope	
FT		/note= "loop region of transmembrane domain"	
FT	Peptide	4148..4219	
FT		/label= epitope	
FT		/note= "loop region of transmembrane domain"	
FT	Peptide	4220..4302	
FT		/label= epitope	
FT		/note= "loop region of transmembrane domain"	
XX			
XX	W0200024780-A2.		
PN			
XX			
PD	04-MAY-2000.		
XX			
PF	25-OCT-1999;	99W0-US25091.	
XX			
XX	26-OCT-1998;	98US-0105731.	
PR	27-OCT-1998;	98US-0105876.	
XX	25-JUN-1999;	99US-0141175.	
XX			
PA	(GENZ) GENZYME CORP.		
XX			
PI	Ibraghimov-Beskrovnaya O, Van-Dellen K, Petry LR;		
XX			
DR	WPI; 2000-350700/30.		
XX	N-PSDB; AAA28423.		
XX			
PT	New isolated antibody, useful for detecting and treating a polycystin kidney disease, comprising an epitope present in transmembrane domain of polycystin and recognizes polycystin-related polypeptides		
XX			
PS	Claim 4; Page 118-126; 126pp; English.		
XX			
CC	Polycystin is encoded by the polycystic kidney disease 1 (PKD1) gene, which has been located on chromosome 16p13.3.		
CC	Epitopes present in the transmembrane domain of polycystin or polycystin-related proteins can be used to raise antibodies (which are claimed). A method for modulating cell-cell adhesion in a suitable tissue, comprising delivering to the tissue an effective amount of an agent that modulates the binding of polycystin in the tissue is also claimed. The PKD1 proteins and antibodies can be used for detecting a polycystin-related polypeptide in a sample (all claimed). The host cells can be used, inter alia, as repositories of polycystin polynucleotides or as vehicles for the production of polycystin polynucleotides and polypeptide. The polynucleotides and gene delivery vehicles are useful in expression systems for the production of polycystin or polycystin-related polypeptide and are useful as hybridization probes to assay for the presence of polycystin polynucleotides or related sequences in a sample. The polynucleotides are also useful in compositions including vaccines for gene therapy. The antibodies and polypeptides are used in standard diagnostic procedures, especially for detecting a polycystin-related polypeptide or tissue containing the polypeptide by contacting a sample suspected of containing the polypeptide with an antibody.		
XX			
SQ	Sequence 4302 AA;		
Query Match 90.68; Score 29; DB 21; Length 4302;			
Best Local Similarity 83.3%; Pred. No. 2.7e+03;			
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;			

PR 03-JUN-1994; 94US-0253524.
 PA (BGHM) BRIGHAM & WOMENS HOSPITAL.
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Glucksmann S, Reiders S, Schneider M;
 XX
 DR WPI; 1996-049618/05.
 DR N-PSDB; AAT11708.
 XX
 XX
 PT DNA encoding poly-cystic kidney disease gene product - for use in
 PT gene therapy of ADPKD, and in the evaluation of treatment for PKD
 XX
 PS Claim 2; Fig 6; 126pp; English.
 XX
 CC This sequence represents a polycystic kidney disease (PKD1) product
 CC which is associated with autosomal dominant polycystic kidney
 CC disease (ADPKD). Mutations within the PKD1 gene are responsible for
 CC approx. 90% of cases of ADPKD. The coding region of the PKD1 gene
 CC is complex and extensive. It covers approx. 60 kb and contains a
 CC total of 46 exons. It has been localised to within a 750 kb chromosomal
 CC region on the short arm of chromosome 16. Approximately the first two
 CC thirds of the PKD1 gene is duplicated several times in a transcribed
 CC fashion elsewhere in the genome. The PKD1 gene also contains extensive
 CC repeated regions of high GC content. A number of the exons have
 CC alternatively spliced forms giving rise to a number of cDNA clones.
 CC The PKD1 protein contains at least 5 distinct peptide domains which
 CC are likely to be involved in protein-protein and/or protein-
 CC carbohydrate interactions. It also shares amino acid similarity
 CC with a number of extracellular matrix proteins. Antibodies raised
 CC against the PKD1 protein may be used in the detection of mutant PKD1
 CC and, therefore, diagnosis of ADPKD. Fragments of the PKD1 gene may be
 CC used in "mini" gene therapy for the treatment of ADPKD.
 XX
 SQ Sequence 4303 AA;

 Query Match 90.6%; Score 29; DB 17; Length 4303;
 Best Local Similarity 83.3%; Pred. No. 2.7e+03;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

 QY 1 CSVTVG 6
 Db 1270 ctvtvg 1275

 RESULT 9
 AAY95558
 ID AAY95558 standard; Protein; 4303 AA.
 AC AAY95558;
 XX
 DT 10-OCT-2000 (first entry)
 XX
 DE Human polycystin-1.
 XX
 KW PKD-1 gene; polycystin-1; polycystic kidney disease;
 KW human; signal transduction.
 XX
 OS Homo sapiens.
 XX
 PN WO2000040711-A2.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00521.
 XX
 PR 06-JAN-1999; 99US-0115127.
 XX
 PA (CALY) CALIFORNIA INST OF TECHNOLOGY.
 XX
 PI Sternberg FW, Barr MM;
 XX
 XX

 Query Match 90.6%; Score 29; DB 21; Length 4303;
 Best Local Similarity 83.3%; Pred. No. 2.7e+03;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

 QY 1 CSVTVG 6
 Db 1270 ctvtvg 1275

 RESULT 10
 AAR75916
 ID AAR75916 standard; Protein; 4339 AA.
 AC AAR75916;
 XX
 DT 14-APR-1996 (first entry)
 XX
 DE Polycystic kidney disease 1 gene product.
 XX
 KW Autosomal dominant polycystic kidney disease; ADPKD;
 KW polycystic kidney disease 1 gene; PKD1; diagnostic; therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO9518225-A1.
 XX
 PD 06-JUL-1995.
 XX
 PF 23-DEC-1994; 94WO-GB02822.
 XX
 PR 14-JUN-1994; 94GB-0011900.
 PR 24-DEC-1993; 93GB-0026470.
 XX
 PA (MEDI-) MEDICAL RES COUNCIL.
 PA (UYLE-) RIJKSUNIV LEIDEN.
 PA (UYRO-) UNIV ROTTERDAM ERASMUS.
 PA (UYWA-) UNIV WALES COLLEGE OF MEDICINE.
 XX
 PI Breuning MH, Halley DJJ, Harris PC, Hesselting ALW;
 PI Hughes J, Janssen LAJ, Nellist MD, Peral B, Peters DJM;
 PI Roelfsema JH, Sampson J, Ward CJ;
 XX
 DR WPI; 1995-246390/32.
 DR N-PSDB; AAQ91438.
 XX

PT Isolated poly:cystic kidney disease 1 gene and its mutants - useful
 PT for treatment and diagnosis of autosomal dominant poly:cystic kidney
 PT disease

PS Disclosure; Fig 10; 119pp; English.

XX A novel protein (AAR75916) is encoded by the polycystic kidney disease
 CC 1 (PKD1) gene (see AAQ91438), which maps to 16p13.3. Mutations at this
 CC locus are associated with autosomal dominant polycystic kidney disease
 CC (ADPKD). The protein can be used to screen actual or suspected
 CC ADPKD patients for normal or mutated PKD1 polypeptide, or is
 CC used to treat or prevent PKD1-associated disorders such as ADPKD by
 CC administration to affected cells.

XX SQ Sequence 4339 AA;

Query Match 90.6%; Score 29; DB 16; Length 4339;
 Best Local Similarity 83.3%; Pred. No. 2.7e+03;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
 I:||||
 Db 1227 ctvtvg 1232

RESULT 11

AAR87539
 ID AAR87539 standard; Protein; 4339 AA.

XX AC AAR87539;

XX DT 02-FEB-1997 (first entry)

XX DE Polycystic kidney disease 1 polypeptide (polycystin).

XX KW Adult polycystic kidney disease; APKD; PKD1 gene; diagnosis;
 XX therapy; polycystin.

XX OS Homo sapiens.

XX PN WO9534649-A2.

XX PD 21-DEC-1995.

XX PF 13-JUN-1995; 95WO-GB01386.

XX PR 14-APR-1995; 95US-0422582.

XX PR 14-JUN-1994; 94GB-0011900.

XX PR 23-DEC-1994; 94WO-GB02822.

XX PR 13-APR-1995; 95GB-0007766.

XX PA (MEDI-) MEDICAL RES COUNCIL.

XX PA (UYLE-) RIJKSUNIV LEIDEN.

XX PA (UYRO-) UNIV ROTTERDAM ERASMUS.

XX PA (UYWA-) UNIV WALES COLLEGE OF MEDICINE.

XX PI Breunig MH, Halley DJJ, Harris PC, Hesselting ALW;

XX PI Hughes J, Janssen LAJ, Nellist MD, Peral B, Peters DJM;

XX PI Roelofsma JH, Sampson J, Ward CJ;

XX DR WPI; 1996-049678/05.

XX DR N-PSDB; AAT08807.

XX PT Isolated polycystic kidney disease I gene and its deletion mutants
 PT - useful in diagnosis and treatment of PKD1-associated disease and
 PT in gene therapy

XX PS Claim 18; Fig 10; 181pp; English.

XX CC PKD1 polypeptide (AAR87539) is encoded by a partial cDNA clone

XX CC (AAT08807) corresponding to the complete human PKD1 gene (see also
 CC AAT13821) apart from its extreme 5' end. This gene is associated

CC with adult polycystic kidney disease (APKD). The polypeptide can
 CC be produced in transformed host cells for use in the diagnosis of
 CC PKD1-associated diseases, for the detection of disease carriers,
 CC and for the treatment or prevention of these diseases. The PKD1
 CC polypeptide plays a role in the suppression or prevention of APKD.

XX SQ Sequence 4339 AA;

Query Match 90.6%; Score 29; DB 17; Length 4339;
 Best Local Similarity 83.3%; Pred. No. 2.7e+03;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
 I:||||
 Db 1227 ctvtvg 1232

RESULT 12

AAM18776

ID AAM18776 standard; Protein; 29 AA.

XX AC AAM18776;

XX DT 12-OCT-2001 (first entry)

XX DE Peptide #5210 encoded by probe for measuring cervical gene expression.

XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
 XX cervical cancer.

XX OS Homo sapiens.

XX PN WO200157278-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00670.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX PI WPI; 2001-488901/53.

XX PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human cervical epithelial cells -

XX PS Claim 27; SEQ ID No 23602; 487pp; English.

XX CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs; see AAI10068-AAI28459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer.

XX CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 29 AA;

[illegible]

XX 31-MAY-2001.
 PD
 XX
 XX 08-NOV-2000; 2000WO-US28888.
 PF
 XX
 XX 24-NOV-1999; 99CN-0124717.
 PR
 XX
 XX (PENY-) PE CORP NY.
 PA
 PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
 PA
 PA (SINO-) SINOGENOMAX CO LTD.
 XX
 PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;
 XX
 DR WPI; 2001-355877/37.
 DR N-PSDB; AAH62817.
 XX
 XX Primary nucleotide sequence of the shrimp white spot Bacilliform virus
 PT (WSBV), useful for producing viral polypeptides that can be used to
 PT screen for agents that are useful for treating WSBV infection -
 XX
 XX Claim 1; Figure 3; 626pp; English.
 PS
 XX
 CC The invention provides the primary nucleotide sequence of the WSBV genome
 CC (AAH62889), predicted transcript sequences (AAH62889-AAH62899) and
 CC encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
 CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
 CC molecules and proteins of the invention are useful for diagnosis and
 CC monitoring viral infection, in screens for antiviral agents and for
 CC monitoring viral gene expression or activity during a treatment regimen.
 CC The nucleic acid molecules are also useful as antisense constructs to
 CC control viral gene expression in infected cells and tissues and to create
 CC transgenic viral resistant shrimp.
 XX
 SQ Sequence 305 AA;

Query Match 87.5%; Score 28; DB 22; Length 305;
 Best Local Similarity 66.7%; Pred. NO. 3.8e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
 ||:|
 Db 23 csmtig 28

Search completed: January 31, 2002, 09:23:08
 Job time: 198 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:24:56 ; Search time 49.78 Seconds
(without alignments)
9.181 Million cell updates/sec

Title: US-09-536-087-11

Perfect score: 32

Sequence: 1 CSVTVG 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	96.9	559	2 A25872	transcription regu
2	29	90.6	104	2 T44890	hypothetical prote
3	29	90.6	134	2 T30680	hypothetical prote
4	29	90.6	138	2 T33647	hypothetical prote
5	29	90.6	357	2 T21152	hypothetical prote
6	29	90.6	392	2 T27303	hypothetical prote
7	29	90.6	410	1 H70342	conserved hypotet
8	29	90.6	440	2 A82369	thiamin-phosphate
9	29	90.6	538	2 H83466	probable ATP-bindi
10	29	90.6	695	2 S70103	probable membrane
11	29	90.6	4302	2 A38971	polycystic kidney
12	28	87.5	48	2 T03817	hypothetical prote
13	28	87.5	143	2 S00718	ribosomal protein
14	28	87.5	158	2 S75349	hypothetical prote
15	28	87.5	220	2 A81911	probable riboflavi
16	28	87.5	239	2 T22581	hypothetical prote
17	28	87.5	234	2 D81106	riboflavin synthas
18	28	87.5	250	2 F70519	hypothetical prote
19	28	87.5	266	1 R5NT2D	ribosomal protein
20	28	87.5	269	2 B39403	probable homocitra
21	28	87.5	273	1 S17442	ribosomal protein
22	28	87.5	274	1 R5NT2	ribosomal protein
23	28	87.5	274	1 R5SP2	ribosomal protein
24	28	87.5	286	1 R5SP2	ribosomal protein
25	28	87.5	300	2 E71534	probable ppp2b met
26	28	87.5	302	2 D81692	conserved hypotet
27	28	87.5	389	2 A43587	major outer membra
28	28	87.5	389	2 I40864	major outer membra
29	28	87.5	389	2 I40739	major outer membra

ALIGNMENTS

RESULT 1

A25872

transcription regulator GRF10 - yeast (Saccharomyces cerevisiae)

N;Alternate names: acid phosphatase synthesis regulatory protein; protein D2350; prot

C;Species: Saccharomyces cerevisiae

C;Date: 16-Aug-1988 #sequence.revision 16-Aug-1988 #text_change 21-Jul-2000

C;Accession: A25872; S67648; S19636; S67412; S72100

R;Sengstag, C.; Hinnen, A.

Nucleic Acids Res. 15, 233-246, 1987

A;Title: The sequence of the Saccharomyces cerevisiae gene PHO2 codes for a regulator

A;Reference number: A25872; MUID:87146334

A;Accession: A25872

A;Molecule type: DNA

A;Residues: 1-559 <SEN>

A;Cross-references: GB:M24613; NID:g4143; PIDN:CAA28729.1; PID:g4144

R;Ballasta, J.P.G.; Remacha, M.; Soler-Mira, A.; Jimenez, A.; Garcia-Cantalejo, J.M.;

submitted to the Protein Sequence Database, July 1996

A;Reference number: S67629

A;Accession: S67648

A;Molecule type: DNA

A;Residues: 1-559 <BAL>

A;Cross-references: EMBL:Z74154; NID:g1431149; PIDN:CAA98673.1; PID:e253225; PID:g143

A;Experimental source: strain S288C

R;McCarthy, B.J.; Creasy, C.L.; Bergman, L.W.

Nucleic Acids Res. 19, 3463, 1991

A;Title: Molecular analysis of a temperature sensitive allele of the PHO2 gene of Sac

A;Reference number: S19636; MUID:91288241

A;Accession: S19636

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 65-156, E', 158-310, N', 312-559 <MCC>

A;Cross-references: EMBL:X34293; NID:g4146; PIDN:CAA38192.1; PID:g4147

A;Experimental source: strain GG100-14D

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1990

R;Boskovic, J.; Saiz, J.E.; Soler-Mira, A.; Garcia-Cantalejo, J.; Revuelta, J.L.; Jim

submitted to the EMBL Data Library, February 1996

A;Reference number: S67406

A;Accession: S67412

A;Molecule type: DNA

A;Residues: 1-559 <BOS>

A;Cross-references: EMBL:X95644; NID:g1199535; PIDN:CAA64906.1; PID:e223187; PID:g119

R;Saiz, J.E.; Buitrago, M.J.; Garcia, R.; Revuelta, J.L.; del Rey, F.

Yeast 12, 1077-1084, 1996

A;Title: The sequence of a 20.3 kb DNA fragment from the left arm of Saccharomyces ce

A;Reference number: S72094; MUID:97051597

A;Accession: S72100

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-559 <SAI>

A;Cross-references: EMBL:X95644; NID:g1199535; PIDN:CAA64906.1; PID:e223187; PID:g119

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1996

C;Genetics:

major outer membra
probable obg - Myc
hypothetical prote
glycine-rich prote
transmission block
lectin - mouse (fr
probable acetyltra
hypothetical prote
circumsporozoite p
ribosomal protein
ribosomal protein
hydrogenase, chain
hypothetical prote
circumsporozoite p

30 28 87.5 389 2 D86577
31 28 87.5 479 2 F70680
32 28 87.5 486 2 B96638
33 28 87.5 608 2 T05442
34 28 87.5 3135 2 A48584
35 27 84.4 139 2 S08576
36 27 84.4 191 2 F83251
37 27 84.4 205 2 C75096
38 27 84.4 225 2 G84180
39 27 84.4 264 2 A44969
40 27 84.4 275 2 T11810
41 27 84.4 276 2 T07531
42 27 84.4 278 2 F75069
43 27 84.4 292 2 C75024
44 27 84.4 298 2 E71200
45 27 84.4 343 2 A29319

A:Gene: SGD:GRF10; PHO2; BAS2
 A:Cross-references: MIPS:YDL106c; SGD:S0002264
 A:Map position: 4L
 C:Function:
 A:Description: required for expression of phosphate pathway and other genes; acts as a co-repressor of PHO2
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:78-134/Domain: homeobox homology <HOX>

Query Match 96.9%; Score 31; DB 2; Length 559;
 Best Local Similarity 83.3%; Pred. No. 41;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
 |||||
 DB 200 CSITVG 205

RESULT 2

T44890
 hypothetical protein MLCB22.15c [Imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 19-May-2000
 C:Accession: T44890

R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z22864

A:Accession: T44890

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-104 <PAR>

A:Cross-references: EMBL:Z98741; PIDN:CABL1380.1

A:Experimental source: cosmid B22

C:Genetics:

A:Note: MLCB22.15c

C:Superfamily: Mycobacterium leprae hypothetical protein MLCB22.15c

Query Match 90.6%; Score 29; DB 2; Length 104;
 Best Local Similarity 83.3%; Pred. No. 25;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
 |||||
 DB 82 CTVTVG 87

RESULT 3

T30680

hypothetical protein 78L - Molluscum contagiosum virus 1

N:Alternate names: MC078L

C:Species: Molluscum contagiosum virus 1

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jul-2000

C:Accession: T30680

R:Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Darai, G.; Moss, B.

Science 273, 813-816, 1996

A:Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host re-

A:Reference number: Z20876; MUID:96325459

A:Accession: T30680

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-134 <SEN>

A:Cross-references: EMBL:U60315; NID:gl491943; PIDN:AAC55206.1; PID:gl492021

C:Genetics:

A:Note: MC078L

C:Superfamily: vaccinia virus F11 protein

Query Match 90.6%; Score 29; DB 2; Length 134;
 Best Local Similarity 83.3%; Pred. No. 31;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6

|||||

DB 289 CTVTVG 294

RESULT 6

T27303

hypothetical protein Y68A4A.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T27303

R:Steward, C.

QY 1 CSVTVG 6
 |||||
 DB 72 CSVTIG 77

RESULT 4

T33647

hypothetical protein K01A2.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T33647

R:Henkhaus, J.; Biewald, T.

submitted to the EMBL Data Library, October 1998

A:Description: The sequence of C. elegans cosmid K01A2.

A:Reference number: Z21381

A:Accession: T33647

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-138 <HEN>

A:Cross-references: EMBL:AF099925; PIDN:AAC69504.1; GSPDB:GN00020; CESP:K01A2.2

A:Experimental source: strain Bristol N2; clone K01A2

C:Genetics:

A:Gene: CESP:K01A2.2

A:Map position: 2

A:Introns: 12/1; 35/3; 88/3

Query Match 90.6%; Score 29; DB 2; Length 138;
 Best Local Similarity 83.3%; Pred. No. 32;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
 |||||
 DB 98 CSLTVG 103

RESULT 5

T21152

hypothetical protein F20E11.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T21152

R:Ainscough, R.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19383

A:Accession: T21152

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-357 <WIL>

A:Cross-references: EMBL:Z81508; PIDN:CAB04138.1; GSPDB:GN00023; CESP:F20E11.7

A:Experimental source: clone F20E11

C:Genetics:

A:Gene: CESP:F20E11.7

A:Map position: 5

A:Introns: 52/3; 88/2; 103/3; 154/3; 190/2; 205/3; 241/2; 270/3; 321/3

Query Match 90.6%; Score 29; DB 2; Length 357;
 Best Local Similarity 83.3%; Pred. No. 75;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
 |||||
 DB 289 CTVTVG 294

submitted to the EMBL Data Library, January 1998

A:Reference number: Z20340
A:Accession: T27303
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-392 <WIL>

A:Cross-references: EMBL:AL021503; PIDN:CAAL6424.1; GSPDB:GN00023; CESP:Y68A4A.10

A:Experimental source: clone Y68A4A

C:Genetics:

A:Gene: CESP:Y68A4A.10

A:Map position: 5

A:Introns: 45/2; 74/3; 139/3; 175/2; 190/3; 241/3; 305/3; 356/3

Query Match

Best Local Similarity 90.6%; Score 29; DB 2; Length 392;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6

I:||||

Db 324 CTVTVG 329

RESULT 7

H70342

conserved hypothetical protein aq_474 - Aquifex aeolicus

C:Species: Aquifex aeolicus

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: H70342

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666

A:Accession: H70342

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-410 <AQF>

A:Cross-references: GB:AE000691; GB:AE000657; NID:g2983118; PIDN:AA06733.1; PID:g298312

A:Experimental source: strain VF5

C:Genetics:

A:Gene: aq_474

C:Superfamily: conserved hypothetical protein b0835

Query Match

Best Local Similarity 90.6%; Score 29; DB 1; Length 410;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6

I:||||

Db 46 CTVTVG 51

RESULT 8

A82369

thiamin-phosphate pyrophosphorylase VC0062 [imported] - Vibrio cholerae (strain N16961)

C:Species: Vibrio cholerae

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: A82369

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;

l. R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: A82369

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-440 <HEI>

A:Cross-references: GB:AE004097; GB:AE003852; NID:g9654454; PIDN:AAF93240.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC0062

A:Map position: 1

Query Match

Best Local Similarity 90.6%; Score 29; DB 2; Length 440;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6

I:||||

Db 92 CTVTVG 97

RESULT 9

H83466

probable ATP-binding component of ABC transporter PA1425 [imported], ~ Pseudomonas aer

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: H83466

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337

A:Accession: H83466

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-538 <STO>

A:Cross-references: GB:AE004572; GB:AE004091; NID:g9947372; PIDN:AA04814.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA1425

Query Match

Best Local Similarity 90.6%; Score 29; DB 2; Length 538;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6

I:||||

Db 397 CAVTVG 402

RESULT 10

S70103

Probable membrane protein YDR338c - Yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae

C>Date: 24-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 29-Oct-1999

C:Accession: S70103

R:Du, Z.

submitted to the EMBL Data Library, March 1996

A:Description: The sequence of S. cerevisiae cosmid 9651.

A:Reference number: S70098

A:Accession: S70103

A:Molecule type: DNA

A:Residues: 1-695 <DUZ>

A:Cross-references: EMBL:U51032; NID:g1230659; PID:g1230665; GSPDB:GN00004; MIPS:YDR3

C:Genetics:

A:Gene: MIPS:YDR338c

A:Map position: 4R

C:Keywords: transmembrane protein

F:339-255/Domain: transmembrane #status predicted <TM1>

F:312-328/Domain: transmembrane #status predicted <TM2>

F:383-399/Domain: transmembrane #status predicted <TM3>

F:412-428/Domain: transmembrane #status predicted <TM4>

F:492-508/Domain: transmembrane #status predicted <TM5>

F:534-550/Domain: transmembrane #status predicted <TM6>

F:563-579/Domain: transmembrane #status predicted <TM7>

F:604-620/Domain: transmembrane #status predicted <TM8>

F:630-646/Domain: transmembrane #status predicted <TM9>

Query Match 90.6%; Score 29; DB 2; Length 695;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CSVTVG 6

|||||

Db 253 CSITVG 258

RESULT 11

T03871

polycystic kidney disease protein 1 precursor - human

C:Species: Homo sapiens (man)

C:Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 08-Oct-1999

C:Accession: A38971; A56520; A44604

R:Harris, P.C.

submitted to GenBank, May 1995

A:Reference number: A38971

A:Accession: A38971

A:Molecule type: mRNA

A:Residues: 1-4302 <HAR>

A:Cross-references: GB:L33243; NID:g904222; PIDN:AAC37576.1; PID:g904223

R:Alexandra Gluecksmann-Kuis, M.; Tayber, O.; Woolf, E.A.; Bougueleret, L.; Deng, N.; Al

Z.; Torosian, S.; Zhou, J.; Reiders, S.T.; Bork, P.; Pohlschmidt, M.; Loehning, C.; Kra

Cell 81, 289-298, 1995

A:Title: Polycystic kidney disease: the complete structure of the PKD1 gene and its pro

A:Reference number: A56520; MUID:95254638

A:Accession: A56520

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-70, 'E', '72-137, 'Q', '139-252, 'A', '254-301, 'D', '303-690, 'P', '692-738, 'R', '740-762,

5-1975, 'V', '1977-3389, 'Q', '3390-3980, 'HV', '3983-4003, 'HV', '4006-4302 <ALE>

A:Cross-references: GB:U24497; NID:g799334; PIDN:AAC50128.1; PID:g799335; GB:U24499

R:Ward, C.J.; Peral, B.; Hughes, J.; Thomas, S.; Gamble, V.; MacCarthy, A.B.; Sloane-Sta

aris, J.J.; Dauwerse, H.G.; Peters, D.J.M.; Breuning, M.H.; Nellist, M.; Brook-Carter, P

W.; van den Ouweland, A.M.W.; Eussen, B.; Verhoef, S.; Lindhout, D.; Halley, D.J.J.

Cell 78, 881-894, 1994

A:Title: The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within

A:Reference number: A44604; MUID:94273192

A:Accession: A44604

A:Status: significant sequence differences

A:Molecule type: mRNA

A:Cross-references: GB:L33243

R:Ward, C.J.; Peral, B.; Hughes, J.; Thomas, S.; Gamble, V.; MacCarthy, A.B.; Sloane-Sta

aris, J.J.; Dauwerse, H.G.; Peters, D.J.M.; Breuning, M.H.; Nellist, M.; Brook-Carter, P

W.; van den Ouweland, A.M.W.; Eussen, B.; Verhoef, S.; Lindhout, D.; Halley, D.J.J.

Cell 78, 724A, 1994

A:Reference number: A38972; MUID:94349375

A:Contents: annotation; erratum

A:Note: This is a revision to the sequence from reference A44604; the citation appears c

R:Ward, C.J.; Peral, B.; Hughes, J.; Thomas, S.; Gamble, V.; MacCarthy, A.B.; Sloane-Sta

aris, J.J.; Dauwerse, H.G.; Peters, D.J.M.; Breuning, M.H.; Nellist, M.; Brook-Carter, P

W.; van den Ouweland, A.M.W.; Eussen, B.; Verhoef, S.; Lindhout, D.; Halley, D.J.J.

Cell 81, 1171, 1995

A:Reference number: A56732

A:Contents: annotation; erratum

A:Note: This is a revision to the sequence from reference A44604

C:Genetics:

A:Gene: GDB:PKD1

A:Cross-references: GDB:120293; OMIM:173900; OMIM:601313

A:Map position: 16p13.3-16p13.3

C:Superfamily: human polycystic kidney disease protein 1; proteoglycan carboxyl-terminal

C:Keywords: duplication

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-4302/Product: polycystic kidney disease protein 1 #status predicted <MAT>

F:123-170/Domain: proteoglycan carboxyl-terminal homology <PCH>

Query Match

Best Local Similarity 90.6%; Score 29; DB 2; Length 4302;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CSVTVG 6

|||||

Db 1270 CSVTVG 1275

RESULT 12

T03817

hypothetical protein (clone NG8) - common tobacco

C:Species: Nicotiana tabacum (common tobacco)

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000

C:Accession: T03817

R:Karrer, E.E.; Beachy, R.N.; Holt, C.A.

Plant Mol. Biol. 36, 681-690, 1998

A:Title: Cloning of tobacco genes that elicit the hypersensitive response.

A:Reference number: Z15096; MUID:98187248

A:Accession: T03817

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-48 <KAR>

A:Cross-references: EMBL:U66273; NID:gl762952; PIDN:AAC49979.1; PID:gl762953

A:Experimental source: clone NG8

Query Match 87.5%; Score 28; DB 2; Length 48;

Best Local Similarity 66.7%; Pred. No. 21;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CSVTVG 6

|||||

Db 26 CSITVG 31

RESULT 13

S00718

ribosomal protein L2 - soybean chloroplast (fragment)

C:Species: chloroplast Glycine max (soybean)

C:Date: 31-Dec-1998 #sequence_revision 31-Dec-1998 #text_change 13-Aug-1999

C:Accession: S00718

R:Spiegelmann, A.; Roux, E.; von Allmen, J.M.; Stutz, E.

Nucleic Acids Res. 16, 1199, 1988

A:Title: The soybean chloroplast genome: complete sequence of the rps19 gene, includi

A:Reference number: S00718; MUID:88143992

A:Accession: S00718

A:Molecule type: DNA

A:Residues: 1-143 <SPT>

A:Cross-references: EMBL:X06429; NID:g14311; PIDN:CAA29735.1; PID:g669076

C:Genetics:

A:Gene: rpl2

A:Genome: chloroplast

C:Superfamily: Escherichia coli ribosomal protein L2

C:Keywords: chloroplast; protein biosynthesis; ribosome

Query Match 87.5%; Score 28; DB 2; Length 143;

Best Local Similarity 83.3%; Pred. No. 55;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CSVTVG 6

|||||

Db 55 CSATVG 60

RESULT 14

S75349

hypothetical protein slr2016 - Synecocystis sp. (strain PCC 6803)

C:Species: Synecocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S75349

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,

O. K.; Okumura, S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocys

s.

A;Reference number: S74322; MUID:97061201
A;Accession: S75349
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-158 <KAN>
A;Cross-references: EMBL:D90904; GB:AB001339; NID:g1652225; PIDN:BAAL7263.1; PID:d101799
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 87.5%; Score 28; DB 2; Length 158;
Best Local Similarity 83.3%; Pred. No. 60;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CSVTVG 6
|||
Db 119 CSTTVG 124

RESULT 15
A81911
probable riboflavin synthase (EC 2.5.1.9) alpha chain NMA1416 [imported] - Neisseria men
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: A81911
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A;Reference number: A81775; MUID:20222556
A;Accession: A81911
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-220 <PAR>
A;Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84656.1; PID:g738007
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: ribC; NMA1416
C;Superfamily: riboflavin synthase alpha chain
C;Keywords: transferase

Query Match 87.5%; Score 28; DB 2; Length 220;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CSVTVG 6
|||
Db 160 CSLTVG 165

Search completed: January 31, 2002, 09:24:57
Job time: 262 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

On protein - protein search, using sw model

Run on: January 31, 2002, 09:20:15 ; Search time 43.81 Seconds
(without alignments)
8.219 Million cell updates/sec

Title: US-09-536-087-6

Perfect score: 84

Sequence: 1 RESHFRGLLNQVHLVF 16

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents_AA:*
- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
 - 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
 - 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
 - 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
 - 5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
 - 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	84	100.0	1172	1	US-08-313-288B-19
2	44	52.4	1170	1	US-08-313-288B-20
3	39	46.4	410	4	US-08-411-760-14
4	39	46.4	1720	2	US-08-477-451-12
5	38	45.2	393	2	US-08-559-303B-74
6	38	45.2	393	4	US-09-175-828-74
7	38	45.2	592	2	US-08-366-490-2
8	38	45.2	592	3	US-08-860-483A-2
9	38	45.2	1417	2	US-08-559-303B-78
10	38	45.2	1417	3	US-08-781-891-78
11	38	45.2	1417	4	US-09-175-828-78
12	36	42.9	243	2	US-08-685-992-35
13	36	42.9	243	2	US-09-144-925-35
14	36	42.9	912	4	US-08-943-768-2
15	36	42.9	2296	2	US-08-286-819A-27
16	36	42.9	2296	3	US-08-980-357-27
17	35	41.7	163	2	US-08-469-537A-83
18	35	41.7	190	2	US-08-469-537A-35
19	35	41.7	239	4	US-08-896-933-21
20	35	41.7	239	4	US-08-896-933-26
21	35	41.7	255	1	US-08-446-918A-2
22	35	41.7	255	2	US-08-580-806-2
23	35	41.7	279	2	US-08-469-537A-51
24	35	41.7	285	2	US-08-469-537A-73
25	35	41.7	449	3	US-08-680-506-7
26	35	41.7	557	4	US-09-027-064-2
27	35	41.7	557	4	US-09-271-815-2

28	35	41.7	635	2	US-08-797-366-5	Sequence 5, Appli
29	35	41.7	635	2	US-08-956-268-5	Sequence 5, Appli
30	35	41.7	816	2	US-08-820-170A-37	Sequence 37, Appl
31	35	41.7	816	3	US-09-055-699-37	Sequence 37, Appl
32	35	41.7	816	4	US-09-273-565-37	Sequence 37, Appl
33	35	41.7	1019	1	US-08-271-364A-7	Sequence 7, Appli
34	35	41.7	1019	2	US-08-222-715B-26	Sequence 26, Appl
35	35	41.7	1727	2	US-08-477-451-10	Sequence 10, Appl
36	34	40.5	89	4	US-08-905-223-433	Sequence 433, App
37	34	40.5	186	1	US-08-026-758-23	Sequence 23, Appl
38	34	40.5	218	4	US-08-811-583-3	Sequence 3, Appli
39	34	40.5	239	4	US-08-896-933-27	Sequence 27, Appl
40	34	40.5	344	2	US-09-055-097-3	Sequence 3, Appli
41	34	40.5	515	2	US-09-073-362-3	Sequence 3, Appli
42	34	40.5	515	2	US-09-243-920-3	Sequence 3, Appli
43	34	40.5	522	1	US-08-639-237-2	Sequence 2, Appli
44	34	40.5	522	1	US-08-975-405-2	Sequence 2, Appli
45	34	40.5	569	1	US-08-306-231-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-313-288B-19
; Sequence 19, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM: Floppy disk
; MEDIUM TYPE: IBM PC compatible
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/313,288B
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1172 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-313-288B-19

Query Match 100.0%; Score 84; DB 1; Length 1172;
Best Local Similarity 100.0%; Pred. No. 9.8e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RESHFRGLLNQVHLVF 16
|||||
DB 198 RESHFRGLLNQVHLVF 213
|||||

RESULT 2
US-08-313-288B-20
; Sequence 20, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313.288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1170 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-313-288B-20

Query Match 52.4%; Score 44; DB 1; Length 1170;
Best Local Similarity 61.5%; Pred. No. 12;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 4 HFRGLQNVLVF 16
:|:|:|:|:|
Db 207 NFGVLQNVRFV 219

RESULT 3
US-08-411-760-14
; Sequence 14, Application US/08411760
; Patent No. 6180373
; GENERAL INFORMATION:
; APPLICANT: WICH, G nter, LEIN-FELDER, Walfred, and
; APPLICANT: BACKMAN, Keith
; TITLE OF INVENTION: Microorganisms for the
; TITLE OF INVENTION: Production of tryptophan and process for the
; TITLE OF INVENTION: Producing the same
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Collard & Roe, P.C.
; STREET: 1077 No. 6180373thrn Boulevard
; CITY: Roslyn
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11576
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect Version 5.1
; SOFTWARE: for DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411.760
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P 42 32 468.8
; FILING DATE: 28 SEPTEMBER 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/02588
; FILING DATE: 23 SEPTEMBER 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Collard, Allison C.
; REGISTRATION NUMBER: 22,532
; REFERENCE/DOCKET NUMBER: SCHMID-PCT
; ATTORNEY/AGENT INFORMATION:
; NAME: Freedman, Edward R.
; REGISTRATION NUMBER: 26,048
; REFERENCE/DOCKET NUMBER: SCHMID-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 365-9802
; TELEFAX: (516) 365-9805
; TELEX: 261176 CRG(UR)
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 410 amino acids
; TYPE: Amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-411-760-14

Query Match 46.4%; Score 39; DB 4; Length 410;
Best Local Similarity 50.0%; Pred. No. 28;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RESHFRGLQNVL 14
:|:|:|:|:|
Db 52 RDAHFIQLRSRTHL 65

RESULT 4
US-08-477-451-12
; Sequence 12, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477.451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113

REFERENCE/DOCKET NUMBER: 0335.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1720 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-451-12

Query Match 46.4%; Score 39; DB 2; Length 1720;
Best Local Similarity 53.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 HFRGLQNVLVF 16
||| | || |
DB 755 HYRWLKKHLSF 767

RESULT 5
US-08-559-303B-74
; Sequence 74, Application US/08559303B
; Patent No. 5824501
; GENERAL INFORMATION:
; APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
; APPLICANT: GRODEN
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: OF BLOOM'S SYNDROME
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/559,303B
; FILING DATE: NOVEMBER 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 63475/65
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 393
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE:
; DESCRIPTION: OTHER NUCLEIC ACID
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
US-08-559-303B-74

Query Match 45.2%; Score 38; DB 2; Length 393;
Best Local Similarity 53.8%; Pred. No. 41;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 RESHFRGLQNWH 13
||| | || |
DB 368 RETHFNLYSMVH 380

RESULT 6
US-09-175-828-74
; Sequence 74, Application US/09175828
; Patent No. 6221643
; GENERAL INFORMATION:
; APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
; APPLICANT: GRODEN
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: OF BLOOM'S SYNDROME
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/175,828
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/559,303
; FILING DATE: NOVEMBER 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 63475/65
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 393
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE:
; DESCRIPTION: OTHER NUCLEIC ACID
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
US-09-175-828-74

Query Match 45.2%; Score 38; DB 4; Length 393;
Best Local Similarity 53.8%; Pred. No. 41;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 RESHFRGLQNWH 13
||| | || |
DB 368 RETHFNLYSMVH 380

RESULT 7
US-08-366-490-2
; Sequence 2, Application US/08366490
; Patent No. 5877403
; GENERAL INFORMATION:
; APPLICANT: McMaster, J. Russell
; APPLICANT: Boeshore, Maury L.
; APPLICANT: Tricoli, David M.
; APPLICANT: Reynolds, John F.
; APPLICANT: Carney, Kim J.
; TITLE OF INVENTION: PAPAYA RINGSPOT VIRUS PROTEASE GENE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitzpatrick, Cella, Harper, and Scinto
; STREET: 277 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10172-0194
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/366.490
; FILING DATE: 30-DEC-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitzpatrick, Cella, Harper, and Scinto
; REFERENCE/DOCKET NUMBER: 4869
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-2400
; TELEFAX: 212-758-2982
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 592 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-366-490-2

Query Match 45.2%; Score 38; DB 2; Length 592;
Best Local Similarity 63.6%; Pred. No. 64;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 HFRGLQNVL 14
||| |
Db 107 HFRDLTNVHV 117

RESULT 8
US-08-860-483A-2
; Sequence 2, Application US/08860483A
; Patent No. 6046384
; GENERAL INFORMATION:
; APPLICANT: McMaster, J. R.
; APPLICANT: Boeshore, Maury L.
; APPLICANT: Tricoli, David M.
; APPLICANT: Reynolds, John F.
; APPLICANT: Carney, Kim J.
; APPLICANT: Slighton, Jerry L.
; APPLICANT: Gonsalves, Dennis
; TITLE OF INVENTION: Papaya Ringspot Virus Nia Protease Gene
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rocky, Milnamow & Katz
; STREET: 180 N. Stetson Avenue, 2 Prudential Plaza,
; CITY: Suite 4700
; STATE: IL
; COUNTRY: USA

ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860.483A
; FILING DATE: 26-JUN-1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueller, Lisa V.
; REGISTRATION NUMBER: 38,978
; REFERENCE/DOCKET NUMBER: SVS3801P0091US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 3126165400
; TELEFAX: 3126165460
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 592 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-860-483A-2

Query Match 45.2%; Score 38; DB 3; Length 592;
Best Local Similarity 63.6%; Pred. No. 64;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 HFRGLQNVL 14
||| |
Db 107 HFRDLTNVHV 117

RESULT 9
US-08-559-303B-78
; Sequence 78, Application US/08559303B
; Patent No. 5824501
; GENERAL INFORMATION:
; APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
; APPLICANT: GRODEN
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: OF BLOOM'S SYNDROME
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/559.303B
; FILING DATE: NOVEMBER 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 63475/65
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1417
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE

; TOPOLOGY: LINEAR
; MOLECULE TYPE:
; DESCRIPTION: OTHER NUCLEIC ACID
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
US-08-559-303B-78

Query Match 45.2%; Score 38; DB 2; Length 1417;
Best Local Similarity 53.8%; Pred. No. 1.7e+02;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 RESHFRGLQNVH 13
||:| | |

Db 1016 RETHFNLYSMVH 1028

RESULT 10
US-08-781-891-78
; Sequence 78, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620tenburg ph.d., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1417 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-781-891-78

Query Match 45.2%; Score 38; DB 3; Length 1417;
Best Local Similarity 53.8%; Pred. No. 1.7e+02;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 RESHFRGLQNVH 13
||:| | |

Db 1016 RETHFNLYSMVH 1028

RESULT 11
US-09-175-828-78
; Sequence 78, Application US/09175828
; Patent No. 6221643
; GENERAL INFORMATION:
; APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
; APPLICANT: GRODEN
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: OF BLOOM'S SYNDROME
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/175,828
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/559,303
; FILING DATE: NOVEMBER 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 63475/65
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1417
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE:
; DESCRIPTION: OTHER NUCLEIC ACID
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
US-09-175-828-78

Query Match 45.2%; Score 38; DB 4; Length 1417;
Best Local Similarity 53.8%; Pred. No. 1.7e+02;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 RESHFRGLQNVH 13
||:| | |

Db 1016 RETHFNLYSMVH 1028

RESULT 12
US-08-685-992-35
; Sequence 35, Application US/08685992
; Patent No. 5912138
; GENERAL INFORMATION:
; APPLICANT: Tonks, Nicholas
; APPLICANT: Flint, Andrew J.

;
; TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASES
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA: US/08/685,992
; FILING DATE: 25-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: CSHL96-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 781-861-6240
; TELEFAX: 781-861-9540
; TELEX:
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-685-992-35

Query Match 42.9%; Score 36; DB 2; Length 243;
Best Local Similarity 70.0%; Pred. No. 53;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ESHFRGLLQN 11
| | | | | : |
Db 55 ESHFRMLAEN 64

RESULT 13
US-09-144-925-35
; Sequence 35, Application US/09144925
; Patent No. 5951979
; GENERAL INFORMATION:
; APPLICANT: Tonks, Nicholas
; APPLICANT: Flint, Andrew J.
; TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASES
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02421-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA: US/09/144,925

;
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/685,992
; FILING DATE: July 25, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: CSHL96-03Z
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 781-861-6240
; TELEFAX: 781-861-9540
; TELEX:
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-144-925-35

Query Match 42.9%; Score 36; DB 2; Length 243;
Best Local Similarity 70.0%; Pred. No. 53;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ESHFRGLLQN 11
| | | | | : |
Db 55 ESHFRMLAEN 64

RESULT 14
US-08-943-768-2
; Sequence 2, Application US/08943768C
; Patent No. 6238881
; GENERAL INFORMATION:
; APPLICANT: Hart, Matthew J.
; TITLE OF INVENTION: No. 6238881el Nucleic Acids and Polypeptides Related to a
; TITLE OF INVENTION: Guanine Exchange Factor of RHO GTPase
; FILE REFERENCE: 1023-US
; CURRENT APPLICATION NUMBER: US/08/943,768C
; CURRENT FILING DATE: 1997-10-06
; EARLIER APPLICATION NUMBER: 60/029,979
; EARLIER FILING DATE: 1996-11-06
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 912
; TYPE: PRT
; ORGANISM: Human pl15 GEF-Rho
; US-08-943-768-2

Query Match 42.9%; Score 36; DB 4; Length 912;
Best Local Similarity 50.0%; Pred. No. 2.4e+02;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RESHFRGLLQNVHLVF 16
| : | | | | : |
Db 55 RPAHLMALLOHVALQF 70

RESULT 15
US-08-286-819A-27
; Sequence 27, Application US/08286819A
; Patent No. 5871910
; GENERAL INFORMATION:
; APPLICANT: ARTHUR, MICHEL
; APPLICANT: DUKTA-MALEN, SYLVIE
; APPLICANT: MOLINAS, CATHERINE
; APPLICANT: COURVALIN, PATRICE
; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE

;; TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPOLYMERIDES, IN PARTICULAR
;; TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR
;; TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS
;; NUMBER OF SEQUENCES: 54
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;; ADDRESSEE: P.C.
;; STREET: 1755 S. Jefferson Davis Highway, Suite 400
;; CITY: Arlington
;; STATE: Virginia
;; COUNTRY: U.S.A.
;; ZIP: 22202
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/286,819A
;; FILING DATE: 05-AUG-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/174,682
;; FILING DATE: 28-DEC-1993
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/917,146
;; FILING DATE: 10-AUG-1992
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/FR/91/00855
;; FILING DATE: 29-OCT-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: FR 9013579
;; FILING DATE: 31-OCT-1990
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Oblon, No. 5871910man F.
;; REGISTRATION NUMBER: 24,618
;; REFERENCE/DOCKET NUMBER: 660-060-0 PCT
;; TELEPHONE: (703) 413-3000
;; TELEFAX: (703) 413-2220
;; TELEX: 248855 OPAT UR
;; INFORMATION FOR SEQ ID NO: 27:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2296 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-286-819A-27

Query Match 42.9%; Score 36; DB 2; Length 2296;
Best Local Similarity 52.9%; Pred. No. 6.6e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

QY 1 RES--HFRGLLNQVHLV 15
||| || :| |||
DB 1334 RESKEHFLSILGAVHLI 1350

Search completed: January 31, 2002, 09:23:55
Job time: 220 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:23:57 ; Search time 43.81 Seconds
(without alignments)
3.596 Million cell updates/sec

Title: US-09-536-087-10
Perfect score: 53
Sequence: 1 WSPWAEW 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	100.0	1172	1	US-08-313-288B-19
2	50	94.3	23	1	US-07-846-531D-5
3	50	94.3	23	2	US-08-488-273-5
4	50	94.3	23	6	5426100-5
5	50	94.3	218	3	US-08-985-526-1
6	50	94.3	239	5	PCT-US93-01652-1
7	50	94.3	441	3	US-08-985-526-3
8	50	94.3	1170	1	US-08-313-288B-20
9	48	90.6	788	2	US-08-918-914-4
10	45	84.9	943	2	US-08-808-982-7
11	45	84.9	943	4	US-09-306-902A-7
12	44	83.0	9	4	US-08-476-134A-34
13	44	83.0	9	6	5190920-7
14	44	83.0	9	6	5506208-8
15	44	83.0	1184	2	US-08-918-914-1
16	44	83.0	1184	3	US-08-996-083-3
17	41	77.4	898	2	US-08-808-982-5
18	41	77.4	898	4	US-09-306-902A-5
19	40	75.5	23	1	US-07-846-531D-7
20	40	75.5	23	2	US-08-488-273-7
21	40	75.5	23	6	5426100-7
22	39	73.6	12	1	US-07-801-812A-1
23	39	73.6	12	1	US-08-487-568-1
24	39	73.6	51	2	US-08-799-173A-11
25	39	73.6	557	1	US-08-313-288B-16
26	39	73.6	560	2	US-08-559-492-5
27	39	73.6	561	2	US-08-559-492-12

28	39	73.6	568	1	US-07-862-021B-14	Sequence 14, Appl
29	39	73.6	568	5	PCT-US93-03164-14	Sequence 14, Appl
30	39	73.6	802	1	US-07-862-021B-12	Sequence 12, Appl
31	39	73.6	802	1	US-08-313-288B-12	Sequence 12, Appl
32	39	73.6	802	5	PCT-US93-03164-12	Sequence 12, Appl
33	39	73.6	807	1	US-07-862-021B-10	Sequence 10, Appl
34	39	73.6	807	1	US-08-313-288B-10	Sequence 10, Appl
35	39	73.6	807	5	PCT-US93-03164-10	Sequence 10, Appl
36	38	71.7	126	3	US-08-513-974B-350	Sequence 350, App
37	38	71.7	818	2	US-08-410-784B-2	Sequence 2, Appl
38	38	71.7	818	4	US-09-346-237-11	Sequence 11, Appl
39	38	71.7	2152	4	US-09-036-987A-3	Sequence 3, Appl
40	38	71.7	2152	4	US-09-370-700-3	Sequence 3, Appl
41	37	69.8	15	4	US-08-702-054B-6	Sequence 6, Appl
42	37	69.8	15	4	US-08-702-054B-8	Sequence 8, Appl
43	37	69.8	16	4	US-08-702-054B-22	Sequence 22, Appl
44	37	69.8	459	1	US-08-313-288B-15	Sequence 15, Appl
45	37	69.8	3724	2	US-08-804-227C-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-313-288B-19
; Sequence 19, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; NUMBER OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1172 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-313-288B-19

Query Match 100.0%; Score 53; DB 1; Length 1172;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WSPWAEW 7
|||||||
Db 384 WSPWAEW 390

RESULT 2
US-07-646-531D-5
; Sequence 5, Application US/07646531D
; Patent No. 5200397
; GENERAL INFORMATION:
; APPLICANT: Deutch, Alan Howard
; APPLICANT: Tuszyński, George Paul
; TITLE OF INVENTION: Peptide Fragments and Analogs of
; TITLE OF INVENTION: Thrombospondin
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. R. Grace & Co.-Conn.
; STREET: 7379 Route 32
; CITY: Columbia
; STATE: Maryland
; COUNTRY: USA
; ZIP: 21044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/646,531D
; FILING DATE: 19910131
; ATTORNEY/AGENT INFORMATION:
; NAME: Appleby, Vanessa L.
; REGISTRATION NUMBER: 33223
; REFERENCE/DOCKET NUMBER: 01-7896
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 531-4515
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-646-531D-5

Query Match 94.3%; Score 50; DB 1; Length 23;
Best Local Similarity 85.7%; Pred. No. 0.14;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
Db 1 WSPWSEW 7

RESULT 3
US-08-488-273-5
; Sequence 5, Application US/08488273
; Patent No. 5840692
; GENERAL INFORMATION:
; APPLICANT: Deutch, Alan H.
; APPLICANT: Tuszyński, George P.
; TITLE OF INVENTION: PEPTIDE FRAGMENTS AND ANALOGS OF
; TITLE OF INVENTION: THROMBOSPONDIN
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,273
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/359,263
; FILING DATE: 19-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/110,146
; FILING DATE: 20-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/896,527
; FILING DATE: 09-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/483,527
; FILING DATE: 22-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9598-2U4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-488-273-5

Query Match 94.3%; Score 50; DB 2; Length 23;
Best Local Similarity 85.7%; Pred. No. 0.14;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
Db 1 WSPWSEW 7

RESULT 4
5426100-5
; Patent No. 5426100
; APPLICANT: DEUTCH, ALAN H.; TUSZYNSKI, GEORGE
; TITLE OF INVENTION: PEPTIDE FRAGMENTS AND ANALOGS OF
; THROMBOSPONDIN
; NUMBER OF SEQUENCES: 12
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,146
; FILING DATE: 20-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 896,527
; FILING DATE: 09-JUN-1992
; APPLICATION NUMBER: 483,527
; FILING DATE: 22-FEB-1990
; SEQ ID NO: 5:
; LENGTH: 23
5426100-5

Query Match 94.3%; Score 50; DB 6; Length 23;
Best Local Similarity 85.7%; Pred. No. 0.14;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
Db 1 WSPWSEW 7

RESULT 5


```

US-08-985-526-1
; Sequence 1, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-985-526-1

Query Match 94.3%; Score 50; DB 3; Length 218;
Best Local Similarity 85.7%; Pred. No. 1.4;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
Db 82 WSPWSEW 88

RESULT 6
PCT-US93-01652-1
; Sequence 1, Application PC/TUS9301652
; GENERAL INFORMATION:
; APPLICANT: Bouck, Noel P.
; APPLICANT: Polverini, Peter J.
; APPLICANT: Good, Deborah J.
; APPLICANT: Frazier, William A.
; TITLE OF INVENTION: Method and Composition for
; TITLE OF INVENTION: Inhibiting Angiogenesis
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut
; STREET: 100 South Wacker Drive, Suite 960
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-4002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/01652
; FILING DATE: 19930222
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/841,656
; FILING DATE: 24-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/464,369
; FILING DATE: 12-JAN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Fentress, Susan B.
; REGISTRATION NUMBER: 31,327
; REFERENCE/DOCKET NUMBER: 92005-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312)-456-8000
; TELEFAX: (312)-456-7776
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 239 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
PCT-US93-01652-1

Query Match 94.3%; Score 50; DB 5; Length 239;
Best Local Similarity 85.7%; Pred. No. 1.5;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
Db 71 WSPWSEW 77

RESULT 7
US-08-985-526-3
; Sequence 3, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 441 amino acids

```

;
; TYPE: amino acid
; TOPOLOGY: linear
US-08-985-526-3

Query Match 94.3%; Score 50; DB 3; Length 441;
Best Local Similarity 85.7%; Pred. No. 2.8;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
 ||||:||
Db 305 WSPWSEW 311

RESULT 8
US-08-313-288B-20
; Sequence 20, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELEPHONE: (212) 391-0526
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1170 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-313-288B-20

Query Match 94.3%; Score 50; DB 1; Length 1170;
Best Local Similarity 85.7%; Pred. No. 7.6;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
 ||||:||
Db 382 WSPWSEW 388

RESULT 9
US-08-918-914-4
; Sequence 4, Application US/08918914
; Patent No. 5876963
; GENERAL INFORMATION:
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy

;
; APPLICANT: Lawton, Michael
; APPLICANT: Magna, Holly
; APPLICANT: Yocum, Sue
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/918,914
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:

INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 788 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1070094
US-08-918-914-4

Query Match 90.6%; Score 48; DB 2; Length 788;
Best Local Similarity 85.7%; Pred. No. 9.7;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
 ||||:||
Db 606 WSPWQEW 612

RESULT 10
US-08-808-982-7
; Sequence 7, Application US/08808982
; Patent No. 5939271
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: Leonardo, E. David
; APPLICANT: Hink, Lindsay
; APPLICANT: Masu, Masayuki
; APPLICANT: Kazuko, Keino-Masu
; TITLE OF INVENTION: Netrin Receptors
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/808,982
; APPLICATION NUMBER: US/08/808,982
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC96-217
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; US-08-808-982-7

Query Match 84.9%; Score 45; DB 2; Length 943;
Best Local Similarity 85.7%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
Db 247 WSSWAEW 253

RESULT 11
US-09-306-902A-7
; Sequence 7, Application US/09306902A
; Patent No. 6277585
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; Hink, Lindsay
; Masu, Masayuki
; Kazuko, Keino-Masu
; TITLE OF INVENTION: Netrin Receptors
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/306,902A
; FILING DATE: 07-May-1999
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC96-217
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid

; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-306-902A-7

Query Match 84.9%; Score 45; DB 4; Length 943;
Best Local Similarity 85.7%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
Db 247 WSSWAEW 253

RESULT 12
US-08-476-134A-34
; Sequence 34, Application US/08476134A
; Patent No. 6239110
; GENERAL INFORMATION:
; APPLICANT: EVAL, JACOB
; APPLICANT: HAMILTON, BRUCE K.
; APPLICANT: TUSZYNSKI, GEORGE P.
; TITLE OF INVENTION: SYNTHETIC ANALOGS OF THROMBOSPONDIN AND THERAPEUTIC USE
; FILE OF INVENTION: THEREOF
; FILE REFERENCE: 07206-0009
; CURRENT APPLICATION NUMBER: US/08/476,134A
; CURRENT FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 07/587,197
; PRIOR FILING DATE: 1990-09-24
; PRIOR APPLICATION NUMBER: 07/483,527
; PRIOR FILING DATE: 1990-02-22
; PRIOR APPLICATION NUMBER: 08/450,738
; PRIOR FILING DATE: 1995-05-25
; PRIOR APPLICATION NUMBER: 08/185,614
; PRIOR FILING DATE: 1994-01-24
; PRIOR APPLICATION NUMBER: 08/024,436
; PRIOR FILING DATE: 1993-03-01
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: analog of thrombospondin
US-08-476-134A-34

Query Match 83.0%; Score 44; DB 4; Length 9;
Best Local Similarity 71.4%; Pred. No. 1.6e+05;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
Db 1 WSPWSPW 7

RESULT 13
5190920-7
; Patent No. 5190920
; APPLICANT: EVAL, JACOB; HAMILTON, BRUCE K.; TUSZYNSKI,
; GEORGE P.
; TITLE OF INVENTION: METHOD FOR USING SYNTHETIC ANALOGS
; OF THROMBOSPONDIN FOR INHIBITING METASTASIS ACTIVITY
; NUMBER OF SEQUENCES: 32
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/587,197
; FILING DATE: 24-SEP-1990
; SEQ ID NO: 7:
; LENGTH: 9

5190920-7

Query Match 83.0%; Score 44; DB 6; Length 9;
Best Local Similarity 71.4%; Pred. No. 1.6e+05;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
 ||||: |
Db 1 WSPWSPW 7

RESULT 14
5506208-8
; Patent No. 5506208
; APPLICANT: EYAL, JACOB; HAMILTON, BRUCE K.; TUSZYNSKI,
; GEORGE P.
; TITLE OF INVENTION: METHOD FOR USING SYNTHETIC ANALOGS OF
; THROMBOSPONDIN FOR INHIBITING ANGIOGENESIS ACTIVITY
; NUMBER OF SEQUENCES: 45
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/408,181
; FILING DATE: 22-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 131,565
; FILING DATE: 04-OCT-1993
; APPLICATION NUMBER: 895,764
; FILING DATE: 09-JUN-1992
; APPLICATION NUMBER: 587,197
; FILING DATE: 24-SEP-1990
; SEQ ID NO: 8
; LENGTH: 9
5506208-8

Query Match 83.0%; Score 44; DB 6; Length 9;
Best Local Similarity 71.4%; Pred. No. 1.6e+05;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
 ||||: |
Db 1 WSPWSPW 7

RESULT 15
US-08-918-914-1
; Sequence 1, Application US/08918914
; Patent No. 5876963
; GENERAL INFORMATION:
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy
; APPLICANT: Lawton, Michael
; APPLICANT: Magna, Holly
; APPLICANT: Yocum, Sue
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/918,914
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0369
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1184 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: ???
CLONE: 422069
US-08-918-914-1

Query Match 83.0%; Score 44; DB 2; Length 1184;
Best Local Similarity 71.4%; Pred. No. 54;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
 ||||: |
Db 152 WSPWSPW 158

Search completed: January 31, 2002, 09:23:58
Job time: 223 sec

THIS PAGE BLANK (USPTO)

7

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:23:06 ; Search time 96.28 Seconds
(without alignments)
5.385 Million cell updates/sec

Title: US-09-536-087-10
Perfect score: 53
Sequence: 1 WSPWAEW 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	100.0	7	21	Human thrombospondin
2	53	100.0	50	20	Human METH2 thombo
3	53	100.0	50	22	TSP2 domain #1. H
4	53	100.0	300	21	Human COMP/TSP-2 c
5	53	100.0	1172	21	Human thrombospond
6	53	100.0	1172	21	Human thrombospond
7	50	94.3	23	12	Thrombospondin pep
8	50	94.3	23	20	Thrombospondin syn
9	50	94.3	50	20	Human METH1 thombo
10	50	94.3	50	22	TSP1 domain #1. H
11	50	94.3	157	21	Amino acid sequenc

12	50	94.3	218	19	AAW40287	Human TSP1 protein
13	50	94.3	218	20	AAW06182	Thrombospondin I f
14	50	94.3	239	14	AAW40823	Human thrombospond
15	50	94.3	441	19	AAW40288	Human concatamers
16	50	94.3	441	20	AAW06183	Thrombospondin I f
17	50	94.3	459	22	AAU02916	Angiotensin conver
18	50	94.3	466	21	AAW43602	Human cancer assoc
19	50	94.3	546	22	AAU02915	Angiotensin conver
20	50	94.3	555	22	AAU02914	Angiotensin conver
21	50	94.3	731	22	AAU02913	Angiotensin conver
22	50	94.3	999	21	AAW94990	Human secreted pro
23	50	94.3	1152	21	AAW00042	Human thrombospond
24	50	94.3	1170	22	AAW90800	Human shear stress
25	50	94.3	1170	22	AAW74450	Human variant thro
26	49	92.5	339	20	AAW60017	Human endometrium
27	45	84.9	943	19	AAW78900	Rat UNC-5 homologu
28	45	84.9	945	22	AAU12244	Human PRO4326 poly
29	44	83.0	9	22	AAW04540	Synthetic peptide
30	44	83.0	56	20	AAW49508	Human METH2 thombo
31	44	83.0	56	22	AAW50010	TSP2 domain #3. H
32	44	83.0	1184	20	AAW74445	Human nucleotide p
33	44	83.0	1184	21	AAW66657	Membrane-bound pro
34	44	83.0	1184	22	AAU12377	Human PRO1188 poly
35	44	83.0	1184	22	AAW65180	Human PRO1188 (UNQ
36	43	81.1	931	22	AAW50691	Human UNC5C protei
37	42	79.2	64	22	AAW06406	Human foetal prote
38	41	77.4	122	22	AAW47165	Flanking region of
39	41	77.4	898	19	AAW78898	Rat UNC-5 homologu
40	40	75.5	23	12	AAW13631	Thrombospondin pep
41	40	75.5	23	20	AAW81484	Thrombospondin syn
42	40	75.5	57	20	AAW49505	Human METH1 thombo
43	40	75.5	57	22	AAW50007	TSP1 domain #3. H
44	40	75.5	242	21	AAW00040	Human COMP/TSP-1 c
45	40	75.5	337	21	AAW32196	Human receptor mol

ALIGNMENTS

RESULT 1
AAW19683
ID AAW19683 standard; Peptide; 7 AA.
XX
AC AAW19683;
XX
DT 05-FEB-2001 (first entry)
XX
Human thrombospondin-2 type 1 repeat peptide 7 (aa384-390).
XX
Thrombospondin-2; TSP-2; human; angiogenesis; cell proliferation;
KW melanoma; tumour; cancer; squamous cell carcinoma; antiangiogenic;
KW prostate cancer; psoriasis; rosecea dermatosis; antitumour;
KW therapy.
XX
Homo sapiens.
OS
XX
WO200057899-A1.
PN
XX
PD 05-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US07835.
XX
PR 31-MAR-1999; 99US-0127221.
XX
PA (GEO) GEN HOSPITAL CORP.
XX
PI Detmar M, Streit M;
XX
DR WPI; 2000-656131/63.
XX
PT Treating a disorder characterized by unwanted cell proliferation e.g.
PT precancerous, cancerous or neoplastic cells or presence of tumor
PT preferably of skin or prostate, comprises increasing thrombospondin-2

PT activity -
 PS Disclosure; Page 40; 73pp; English.
 XX
 XX The present sequence is that of peptide 7, derived from the first
 CC type 1 repeat (amino acids 384-390) of human thrombospondin-2
 CC (TSP-2, see AAB19677). The peptide is 1 of 5 synthetic peptides (see
 CC AAB19679-83) used with human dermal microvascular endothelial cells
 CC (HMEC) to determine the effect of TSP-2 on endothelial cell
 CC migration. Peptide 7 inhibited MDMEC migration by 47.6% in
 CC contrast to peptides derived from the TSP-2 procollagen domain,
 CC revealing an important role of this TSP-2 specific peptide for the
 CC anti-angiogenic activity of TSP-2. The invention is based on the
 CC discovery that overexpression of TSP-2 decreases tumor size in
 CC vivo, and features methods for modulating unwanted angiogenesis and
 CC cell proliferation by increasing TSP-2 activity.
 XX
 SQ Sequence 7 AA;
 Query Match 100.0%; Score 53; DB 21; Length 7;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 WSPWAEW 7
 Db 1 WSPWAEW 7
 RESULT 2
 AAY49506
 ID AAY49506 standard; Protein; 50 AA.
 XX
 AC AAY49506;
 XX
 DT 10-JAN-2000 (first entry)
 XX
 DE Human METH2 thrombospondin-like domain #1.
 XX
 KW Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
 KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
 KW angiogenesis inhibitor; abnormal wound healing; inflammation;
 KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
 KW diabetic retinopathy; macula degeneration; haemangioma; detection;
 KW arterial-venous malformation; immune deficiency.
 XX
 OS Homo sapiens.
 XX
 PN WO9937660-A1.
 XX
 PD 29-JUL-1999.
 XX
 PF 22-JAN-1999; 99WO-US01313.
 XX
 PR 23-JAN-1998; 98US-0072298.
 PR 28-AUG-1998; 98US-0098539.
 XX
 PA (IRUE/) IRUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBE/) RUBEN S M.
 XX
 PI IrueLA-Arispe L, Hastings GA, Ruben SM;
 XX
 DR WPI; 1999-590684/50.
 XX
 PT New isolated metalloprotease thrombospondin polypeptides, useful for
 PT treating hyperproliferative disorders, cancers or autoimmune disorders
 PT
 PS Disclosure; Fig 5; 457pp; English.
 XX
 CC AAY32000 and AAY32001 encode, and AAY49501 and AAY49502 represent, human
 CC metalloprotease thrombospondin (METH) proteins METH1 and METH2

CC respectively. METH1 and METH2 have been found to be potent inhibitors of
 CC angiogenesis both in vitro and in vivo. They can be used for treating
 CC cancer and other disorders related to angiogenesis including abnormal
 CC wound healing, inflammation, rheumatoid arthritis, psoriasis,
 CC endometrial bleeding disorders, diabetic retinopathy, some forms of
 CC macula degeneration, haemangiomas, and arterial-venous malformations.
 CC They may be useful in treating deficiencies or disorders of the immune
 CC system, by activating or inhibiting the proliferation, differentiation,
 CC or mobilisation (chemotaxis) of immune cells. The etiology of these
 CC immune deficiencies or disorders may be genetic, somatic, such as
 CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
 CC toxins), or infectious. They can also be used to treat inflammatory
 CC conditions, both chronic and acute conditions. The products can also be
 CC used for detection and diagnosis. AAY32002 to AAY32080, and AAY49503 to
 CC AAY49511 represent sequences given in the exemplification of the present
 CC invention.
 XX
 SQ Sequence 50 AA;
 Query Match 100.0%; Score 53; DB 20; Length 50;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 WSPWAEW 7
 Db 4 WSPWAEW 10
 RESULT 3
 AAB50008
 ID AAB50008 standard; Protein; 50 AA.
 XX
 AC AAB50008;
 XX
 DT 19-MAR-2001 (first entry)
 XX
 DE TSP2 domain #1.
 XX
 KW Human; METH; metalloprotease; thrombospondin; angiogenesis inhibition;
 KW cancer therapy; benign tumour; ocular angiogenic disease;
 KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
 KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
 KW scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
 KW coronary collateral; cerebral collateral; arteriovenous malformation;
 KW ischaemic limb angiogenesis; Osler-Webber syndrome; TSP domain;
 KW plaque neovascularisation; telangiectasia; haemophilic joint;
 KW angiofibroma; fibromuscular dysplasia; wound granulation;
 KW Crohn's disease; atherosclerosis; birth control.
 XX
 OS Homo sapiens.
 XX
 PN WO200071577-A1.
 XX
 PD 30-NOV-2000.
 XX
 PF 25-MAY-2000; 2000WO-US14462.
 XX
 PR 25-MAY-1999; 99US-0318208.
 PR 20-JUL-1999; 99US-0144882.
 PR 10-AUG-1999; 99US-0147823.
 PR 13-AUG-1999; 99US-0373658.
 PR 22-DEC-1999; 99US-0171503.
 PR 22-FEB-2000; 2000US-0183792.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA (IRUE/) IRUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBE/) RUBEN S M.
 PA (JONA/) JONAK Z L.
 PA (TRUL/) TRULLI S H.

PA (FORN/) FORNWALD J A.
PA (TERR/) TERRETT J A.
PI Iruela-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trullis SH;
PI Fornwald JA, Terrett JA;
XX WPI; 2001-025136/03.
DR
XX METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT inhibit angiogenesis in the treatment of disorders such as cancer,
PT rheumatoid arthritis and psoriasis -
XX Disclosure; Fig 5; 768pp; English.
XX
XX The present invention relates to human METH1 and METH2, (ME for
CC metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
CC The present sequence is a TSP domain which are found in METH protein.
CC METH can be used for inhibiting angiogenesis in an individual, and for
CC treating cancer, benign tumours, an ocular angiogenic disease,
CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
CC vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
CC fibromuscular dysplasia, wound granulation, Crohn's disease or
CC atherosclerosis. METH can also be used in birth control. METH can also
CC be used in diagnostic methods for the prognosis of cancer.
XX
SQ Sequence 50 AA;

Query Match 100.0%; Score 53; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.16; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 1 WSPWAEW 7
Db 4 wspwaw 10
|||||

RESULT 4
AAB00041
ID AAB00041 standard; Protein; 300 AA.
XX
AC AAB00041;
XX
DT 08-NOV-2000 (first entry)
XX
DE Human COMP/TSP-2 chimeric protein.
XX
KW TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
KW thrombospondin; angiogenesis; tumour; treatment; cancer;
KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
KW glaucoma.
XX
OS Homo sapiens.
OS Synthetic.
XX
XX WO200044908-A2.
XX
XX 03-AUG-2000.
XX
XX 01-FEB-2000; 2000WO-US02482.
XX
XX 01-FEB-1999; 99US-0118053.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
XX Lawler JW;
XX
XX WPI; 2000-514823/46.
DR N-PSDB; AAA47735.
DR

XX Nucleic acids encoding chimeric proteins such as cartilage oligomeric
PT matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for
PT inhibiting angiogenesis and treating diseases such as cancer
XX Claim 50; Fig 5a-b; 40pp; English.
PS
XX New nucleic acids are described which encode a protein comprising
CC the second and third type 1 repeats of human TSP (thrombospondin)-1,
CC but not the TGF (transforming growth factor)-beta activation region
CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing
CC the second and third type-1 repeats and the COMP (cartilage
CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was
CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1
CC caused inhibition of the growth of tumours in mice models.
CC Thus the nucleic acids and proteins may be useful for treating
CC angiogenesis related diseases such as cancer (by reducing the rate of
CC growth and size of tumours), arthritis, psoriasis, diabetic
CC retinopathy, corneal graft rejection, and glaucoma. They may also be
CC used for treating human immunodeficiency virus (HIV) infection.
CC Anti-angiogenic therapy has little toxicity, does not require the
CC therapeutic agent to enter tumour cells or cross the blood-brain
CC barrier, controls tumour growth independently of growth of
CC tumour cell heterogeneity, and does not induce drug resistance.
XX
SQ Sequence 300 AA;

Query Match 100.0%; Score 53; DB 21; Length 300;
Best Local Similarity 100.0%; Pred. No. 0.97; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 1 WSPWAEW 7
Db 132 wspwaw 138
|||||

RESULT 5
AAB19677
ID AAB19677 standard; Protein; 1172 AA.
XX
AC AAB19677;
XX
DT 05-FEB-2001 (first entry)
XX
DE Human thrombospondin-2.
XX
KW Thrombospondin-2; TSP-2; human; angiogenesis; cell proliferation;
KW melanoma; tumour; cancer; squamous cell carcinoma; antiangiogenic;
KW prostate cancer; psoriasis; rosecea dermatosis; antitumour;
KW therapy.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX 382..429
XX /note= "type 1 repeat"
XX 384..390
XX /note= "WSPWAEW sequence involved in antiangiogenic
XX activity"
XX 438..490
XX /note= "type 1 repeat"
XX 495..547
XX /note= "type 1 repeat"
XX
XX WO200057899-A1.
XX
XX 05-OCT-2000.
XX
XX 24-MAR-2000; 2000WO-US07835.
XX
XX 31-MAR-1999; 99US-0127221.
XX

PA (GEHO) GEN HOSPITAL CORP.

XX PI Detmar M, Streit M;

XX DR WPI: 2000-656131/63.

DR N-PSDB; AAA88669.

XX Treating a disorder characterized by unwanted cell proliferation e.g.
PT precancerous, cancerous or neoplastic cells or presence of tumor
PT preferably of skin or prostate, comprises increasing thrombospondin-2
PT activity

XX PS Disclosure; Fig 2; 73pp; English.

XX The present sequence is that of human thrombospondin-2 (TSP-2).
CC The invention is based on the discovery that overexpression of
CC TSP-2 decreases tumor size in vivo, and features methods for
CC modulating unwanted angiogenesis and tumor growth. Treatment of
CC unwanted cell proliferation or angiogenesis involves increasing
CC TSP-2 activity. This is achieved by administering an agent which
CC increases TSP-2 activity, especially a TSP-2 polypeptide, a TSP-2
CC derived polypeptide or retro-inverso peptide, a nucleic acid
CC encoding TSP-2, an agonist of TSP-2, or an agent that increases
CC TSP-2 gene expression. The TSP-2 polypeptide may include at least
CC one type 1 repeat such as the WSPWAEW peptide (see AAB19683),
CC which is involved in the antiangiogenic activity of TSP-2. The
CC method is used to treat a disorder characterised by pre-cancerous,
CC cancerous or neoplastic cells, or the presence of a tumour, or a
CC disorder that affects epithelial tissues resulting in unwanted
CC skin cell proliferation. Such disorders include malignant
CC melanoma, prostate cancer, squamous cell carcinoma, aged skin,
CC rosacea dermatosis, psoriasis, and skin damage caused by
CC photoradiation (all claimed). Evaluating the presence of TSP-2
CC nucleic acid or protein is useful for diagnosing a subject at risk
CC of unwanted cell proliferation or angiogenesis. Methods are also
CC provided of identifying compounds that modulate TSP-2 activity.

XX SQ Sequence 1172 AA;

Query Match 100.0%; Score 53; DB 21; Length 1172;

Best Local Similarity 100.0%; Pred. No. 3.8;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7

Db 384 wspwaeW 390

RESULT 6

ID AAB00043
XX AAB00043 standard; Protein; 1172 AA.

XX AC AAB00043;

XX DT 08-NOV-2000 (first entry)

XX DE Human thrombospondon-2 (TSP-2).

XX TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
KW thrombospondin; angiogenesis; tumour; treatment; cancer;
KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
KW glaucoma.

XX OS Homo sapiens.

XX Key Location/Qualifiers
FH Region 381..436
FT /label= Type 1 repeat region
FT /label= Type 1 repeat region
FT /label= Type 1 repeat region
FT /label= Type 1 repeat region

XX WO200044908-A2.

XX PD 03-AUG-2000.

XX PF 01-FEB-2000; 2000WO-US02482.

XX PR 01-FEB-1999; 99US-0118053.

XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX PI Lawler JW;

XX DR WPI; 2000-514823/46.

XX Nucleic acids encoding chimeric proteins such as cartilage oligomeric
PT matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for
PT inhibiting angiogenesis and treating diseases such as cancer
XX Disclosure; Fig 2; 40pp; English.

XX New nucleic acids are described which encode a protein comprising
CC the second and third type 1 repeats of human TSP (thrombospondin)-1,
CC but not the TGF (transforming growth factor)-beta activation region
CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing
CC the second and third type-1 repeats and the COMP (cartilage
CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was
CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1
CC caused inhibition of the growth of tumours in mice models.
CC Thus the nucleic acids and proteins may be useful for treating
CC angiogenesis related diseases such as cancer (by reducing the rate of
CC growth and size of tumours), arthritis, psoriasis, diabetic
CC retinopathy, corneal graft rejection, and glaucoma. They may also be
CC used for treating human immunodeficiency virus (HIV) infection.
CC Anti-angiogenic therapy has little toxicity, does not require the
CC therapeutic agent to enter tumour cells or cross the blood-brain
CC barrier, controls tumour growth independently of growth of
CC tumour cell heterogeneity, and does not induce drug resistance.

XX SQ Sequence 1172 AA;

Query Match 100.0%; Score 53; DB 21; Length 1172;

Best Local Similarity 100.0%; Pred. No. 3.8;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7

Db 384 wspwaeW 390

RESULT 7

ID AAR13629
XX AAR13629 standard; peptide; 23 AA.

XX AC AAR13629;

XX DT 31-OCT-1991 (first entry)

XX DE Thrombospondin peptide p11.

XX Antiviral agent; wound healing; platelet aggregation; thrombotic;
KW thrombolytic.

XX OS Synthetic.

XX PN EP443404-A.

XX PD 28-AUG-1991.

XX PF 11-FEB-1991; 91EP-0101908.

XX PR 31-JAN-1991; 91US-0646531.

C 6444 /
13-20

PR 22-FEB-1990; 90US-0483527.
 XX (GRAC) GRACE W R & CO-CONN.
 PA (WEDI-) MED COLLEGE OF PENNSYLVANIA.
 XX Deutch AH, Tuszynski GP;
 XX PI
 XX XX
 DR WPI; 1991-254044/35.
 XX
 XX New peptide fragments and analogues of thrombospondin - useful
 PT for inhibiting tumour metastasis, as clotting agents and to
 PT promote or inhibit cell adhesion and immune modulation.
 XX
 XX Claim 3; Page 26; 30pp; English.
 XX
 CC The peptide is a synthetic fragment of human thrombospondin and is
 CC based on the sequence motif of Robson et al (Nature (1988) 335:79-
 CC 82). It has thrombospondin-like activity and can be used:
 CC (1) to inhibit tumour cell metastasis and atherosclerosis;
 CC (2) to promote or inhibit platelet aggregation, angiogenic activity,
 CC thrombotic or thrombolytic activity, immune modulation and cell
 CC adhesion;
 CC (3) to promote wound healing; and
 CC (4) as an antiviral agent (interferes with cell adhesion).
 CC The peptide is prep. by std. synthesis techniques.
 CC See also AAR13626-R13641.
 XX
 XX Sequence 23 AA;
 SQ

Query Match 94.3%; Score 50; DB 12; Length 23;
 Best Local Similarity 85.7%; Pred. No. 0.2;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
 Db 1 wspwsew 7
 |||||

RESULT 8
 AAW81482
 ID AAW81482 standard; peptide; 23 AA.
 XX AC
 AC AAW81482;
 XX
 DT 29-JAN-1999 (first entry)
 XX
 DE Thrombospondin synthetic analog compound p11.
 XX
 KW Thrombospondin; thrombin sensitive protein; TSP; cell-adhesion;
 KW mitogenic activity; chemotactic; haemostatic; tumour; microbial;
 KW parasite; metastasis; platelet aggregation; fibrinolytic; malaria;
 KW immune modulation; wound healing; atherosclerosis; angiogenesis;
 KW complement modulator; diagnostic reagent.
 XX
 OS Synthetic.
 XX
 XX US5840692-A.
 XX
 PD 24-NOV-1998.
 XX
 XX 07-JUN-1995; 95US-0488273.
 XX
 XX 07-JUN-1995; 95US-0488273.
 PR 22-FEB-1990; 90US-0483527.
 PR 09-JUN-1992; 92US-0896527.
 PR 20-AUG-1993; 93US-0110146.
 PR 19-DEC-1994; 94US-0359263.
 XX
 XX (GRAC) GRACE & CO-CONN W R.
 PA (UYAL-) UNIV ALLEGHENY HEALTH SCI.
 XX
 XX Deutch AH, Tuszynski GP;

XX WPI; 1999-034078/03.
 XX
 PT Method for mimicking or inhibiting thrombospondin activity - using
 PT thrombospondin peptides
 XX
 XX Claim 3; Column 29; 19pp; English.
 XX
 CC Sequences AAW81478 to AAW81491 represent syntehtic peptide analogs of
 CC thrombospondin (thrombin sensitive protein or TSP) that retain
 CC thrombospondin-like activity. The invention provides a method for (a)
 CC mimicking a biological activity of thrombospondin; (b) promoting
 CC thrombotic activity; or (c) inhibiting a biological activity of
 CC thrombospondin other than thrombotic activity. The method comprises
 CC administering any of these peptides. Biological activities of
 CC thrombospondin include cell adhesion-promoting activity; cell mitogenic
 CC activity; cell chemotactic activities; haemostatic activities, and
 CC activities that derive from these activities, e.g. tumour-cell,
 CC microbial or parasite metastasis activity, platelet aggregating
 CC activity, fibrinolytic activity and immune modulation. The peptides are
 CC capable of inhibiting tumour metastasis. The peptides are useful in wound
 CC healing, atherosclerosis, malaria, thrombotic and thrombolytic conditions
 CC and angiogenesis, and as cell attachment promoters, complement modulators
 CC and diagnostic reagents.
 XX
 XX Sequence 23 AA;
 SQ

Query Match 94.3%; Score 50; DB 20; Length 23;
 Best Local Similarity 85.7%; Pred. No. 0.2;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSPWAEW 7
 Db 1 wspwsew 7
 |||||

RESULT 9
 AAY49503
 ID AAY49503 standard; Protein; 50 AA.
 XX AC
 AC AAY49503;
 XX
 DT 10-JAN-2000 (first entry)
 XX
 DE Human METH1 thrombospondin-like domain #1.
 XX
 KW Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
 KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
 KW angiogenesis inhibitor; abnormal wound healing; inflammation;
 KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
 KW diabetic retinopathy; macula degeneration; haemangioma; detection;
 KW arterial-venous malformation; immune deficiency.
 XX
 OS Homo sapiens.
 XX
 XX WO9937660-A1.
 XX
 PD 29-JUL-1999.
 XX
 XX 22-JAN-1999; 99WO-US01313.
 XX
 XX 23-JAN-1998; 98US-0072298.
 PR 28-AUG-1998; 98US-0098539.
 XX
 XX (IRUE/) IRUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBE/) RUBEN S M.
 XX
 XX IrueLa-Arispe L, Hastings GA, Ruben SM;
 FI WPI; 1999-590684/50.
 XX
 DR
 XX

PT New isolated metalloprotease thrombospondin polypeptides, useful for
 PT treating hyperproliferative disorders, cancers or autoimmune disorders
 XX
 XX
 PS Disclosure; Fig 5; 457pp; English.
 XX
 XX AA232000 and AA232001 encode, and AA49501 and AA49502 represent, human
 CC metalloprotease thrombospondin (METH) proteins METH1 and METH2
 CC respectively. METH1 and METH2 have been found to be potent inhibitors of
 CC angiogenesis both in vitro and in vivo. They can be used for treating
 CC cancer and other disorders related to angiogenesis including abnormal
 CC wound healing, inflammation, rheumatoid arthritis, psoriasis,
 CC endometrial bleeding disorders, diabetic retinopathy, some forms of
 CC macula degeneration, haemangiomas, and arterial-venous malformations.
 CC They may be useful in treating deficiencies or disorders of the immune
 CC system, by activating or inhibiting the proliferation, differentiation,
 CC or mobilisation (chemotaxis) of immune cells. The etiology of these
 CC immune deficiencies or disorders may be genetic, somatic, such as
 CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
 CC toxins), or infectious. They can also be used to treat inflammatory
 CC conditions, both chronic and acute conditions. The products can also be
 CC used for detection and diagnosis. AA232002 to AA232080, and AA49503 to
 CC AA49511 represent sequences given in the exemplification of the present
 CC invention.
 XX
 XX Sequence 50 AA;

Query Match 94.3%; Score 50; DB 20; Length 50;
 Best Local Similarity 85.7%; Pred. No. 0.44;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
 ||||:|
 Db 4 wspwsew 10

RESULT 10
 AAB50005
 ID AAB50005 standard; Protein; 50 AA.
 AC AAB50005;
 XX
 XX 19-MAR-2001 (first entry)
 XX
 XX TSPI domain #1.
 XX
 XX Human; METH; metalloprotease; thrombospondin; angiogenesis inhibition;
 KW cancer therapy; benign tumour; ocular angiogenic disease;
 KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
 KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
 KW scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
 KW coronary collateral; cerebral collateral; arteriovenous malformation;
 KW ischaemic limb angiogenesis; Osler-Webber syndrome; TSP domain;
 KW plaque neovascularisation; telangiectasia; haemophilic joint;
 KW angiofibroma; fibromuscular dysplasia; wound granulation;
 KW Crohn's disease; atherosclerosis; birth control.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200071577-A1.
 PN
 XX
 XX 30-NOV-2000.
 PD
 XX
 XX 25-MAY-2000; 2000WO-US14462.
 PF
 XX
 XX 25-MAY-1999; 99US-0318208.
 PR
 XX 20-JUL-1999; 99US-0144882.
 PR
 XX 10-AUG-1999; 99US-0147823.
 PR
 XX 13-AUG-1999; 99US-0373658.
 PR
 XX 22-DEC-1999; 99US-0171503.
 PR
 XX 22-FEB-2000; 2000US-0183792.

PA (HUMA-) HUMAN GENOME SCI INC.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA (IRUE/) IRUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBE/) RUBEN S M.
 PA (JONA/) JONAK Z L.
 PA (TRUL/) TRULLI S H.
 PA (FORN/) FORNWALD J A.
 PA (TERR/) TERRETT J A.
 XX
 PI IrueLA-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
 PI Fornwald JA, Terrett JA;
 XX
 XX WPI; 2001-025136/03.
 XX
 XX METH1 and METH2 polynucleotides and encoded polypeptides, used to
 PT inhibit angiogenesis in the treatment of disorders such as cancer,
 PT rheumatoid arthritis and psoriasis -
 XX
 PS Disclosure; Fig 5; 768pp; English.
 XX
 CC The present invention relates to human METH1 and METH2, (ME for
 CC metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
 CC The present sequence is a TSP domain which are found in METH protein.
 CC METH can be used for inhibiting angiogenesis in an individual, and for
 CC treating cancer, benign tumours, an ocular angiogenic disease,
 CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
 CC vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
 CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
 CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
 CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
 CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
 CC fibromuscular dysplasia, wound granulation, Crohn's disease or
 CC atherosclerosis. METH can also be used in birth control. METH can also
 CC be used in diagnostic methods for the prognosis of cancer.
 XX
 XX Sequence 50 AA;

Query Match 94.3%; Score 50; DB 22; Length 50;
 Best Local Similarity 85.7%; Pred. No. 0.44;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7
 ||||:|
 Db 4 wspwsew 10

RESULT 11
 AAB08133
 ID AAB08133 standard; Protein; 157 AA.
 XX
 XX AAB08133;
 AC
 XX
 XX 04-DEC-2000 (first entry)
 DT
 XX
 XX Amino acid sequence of htSP-1 Type I repeat peptide (TRP).
 DE
 XX
 XX Type 1 repeat; thrombospondin-1; TSP-1; Type I repeat peptide; TRP;
 KW KIAA0688; KIAA0550; angiogenesis inhibitory protein; angiogenesis;
 KW cancer; tumour; rheumatoid arthritis; psoriasis; retinopathy;
 KW ocular angiogenic disease; macular degeneration; corneal graft rejection;
 KW neovascular glaucoma; retrolental fibroplasia; rubeosis;
 KW Osler-Webber Syndrome; myocardial angiogenesis; haemophilic joint;
 KW plaque neovascularisation; telangiectasia; angiofibroma;
 KW wound granulation; coronary collateral; cerebral collateral;
 KW arteriovenous malformation; ischemic limb angiogenesis;
 KW neovascular glaucoma; retrolental fibroplasia; fracture; vasculogenesis;
 KW heliobacter related disease; hematopoiesis; ovulation; menstruation;
 KW placentalation; birth control; embryo implantation.
 XX
 XX Homo sapiens.
 OS

XX WO200047622-A2.
 XX 17-AUG-2000.
 XX 15-FEB-2000; 2000WO-GB00520.
 XX 15-FEB-1999; 99GB-0003408.
 XX (OXFO-) OXFORD BIOMEDICA UK LTD.
 XX Mazarakis N, Martin-Rendon E, Kingsman SM;
 XX WPI; 2000-549137/50.
 XX Non-naturally occurring Type I repeat peptide (TRP) derived from human
 PT thrombospondin-1, KIAA0688 or KIAA0550 proteins, useful in the
 PT treatment of angiogenesis and/or cancer -
 XX Claim 4; Fig 2; 84pp; English.
 XX The present sequence represents the amino acid of human thrombospondin-1
 CC (HTSP-1) Type I repeat peptide (TRP). The specification describes a
 CC non-naturally occurring TRP, derived from HTSP-1, KIAA0688 or KIAA0550
 CC proteins. The TRP protein is an angiogenesis inhibitory protein. TRP
 CC proteins are used for the treatment of conditions associated with
 CC angiogenesis and cancer. Angiogenic mediated diseases include tumours,
 CC rheumatoid arthritis, psoriasis, ocular angiogenic diseases, retinopathy
 CC of prematurity, macular degeneration, corneal graft rejection,
 CC neovascular glaucoma, retrolental fibroplasia, rubeosis), Osler-Webber
 CC Syndrome, myocardial angiogenesis, plaque neovascularisation,
 CC telangiectasia, haemophilic joints, angiofibroma, wound granulation,
 CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
 CC ischemic limb angiogenesis, neovascular glaucoma, retrolental
 CC fibroplasia, heliobacter related diseases, fractures, vasculogenesis,
 CC hematopoiesis, ovulation, menstruation and placentation. TRPs are also
 CC useful in the treatment of disease of excessive or abnormal stimulation
 CC of endothelial cells. TRP can also be used as a birth control agent, as
 CC it prevents the vascularisation required for embryo implantation.
 XX Sequence 157 AA;
 SQ
 Query Match 94.3%; Score 50; DB 21; Length 157;
 Best Local Similarity 85.7%; Pred. No. 1.4;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WSPWAEW 7
 Db 3 wspwsew 9
 RESULT 12
 AAW40287
 ID AAW40287 standard; Protein; 218 AA.
 XX AAW40287;
 XX 18-AUG-1998 (first entry)
 DT Human TSPI protein.
 XX
 DE
 XX
 KW TSPI; thrombospondin; anti-angiogenic; cationic vehicle; gene therapy;
 KW liposome; DNA complex; tumour suppressor protein; treatment; neoplastic;
 KW metabolic disease; tumour.
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 160 /note= "encoded by CGG"
 FT Misc-difference 217 /note= "encoded by CT"
 FT Misc-difference 217 /note= "encoded by CT"

FT Peptide 12..27
 FT /label= anti-angiogenic peptide
 FT Peptide 142..159
 FT /label= anti-angiogenic peptide
 FT Peptide 199..217
 FT /label= anti-angiogenic peptide
 XX EP819758-A2.
 XX PN
 XX 21-JAN-1998.
 PD
 XX 16-JUL-1997; 97EP-0112154.
 PF
 XX 16-JUL-1996; 96US-0680845.
 PR
 XX (MIXS/) MIXSON A J.
 PA
 XX Mixson AJ;
 PI
 XX WPI; 1998-078839/08.
 XX DR
 XX N-PSDB; AAV10493.
 DR
 XX Complexes of DNA encoding anti-angiogenic peptide - with cationic
 PT liposome(s) or cationic polymer, useful for, e.g. gene therapy of
 PT tumours
 PT
 PS Claim 9; Page 6; 47pp; English.
 XX
 CC This sequence represents the thrombospondin gene TSPI which is used
 CC in a method to produce a cationic vehicle consisting of a cationic
 CC liposome:DNA complex where the DNA encodes an anti-angiogenic peptide or
 CC tumour suppressor protein. Such complexes are used for treatment of
 CC neoplastic and metabolic diseases especially for gene therapy of tumours.
 XX Sequence 218 AA;
 SQ
 Query Match 94.3%; Score 50; DB 19; Length 218;
 Best Local Similarity 85.7%; Pred. No. 1.9;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WSPWAEW 7
 Db 82 wspwsew 88
 RESULT 13
 AAY06182
 ID AAY06182 standard; Protein; 218 AA.
 XX AAY06182;
 XX 16-AUG-1999 (first entry)
 DT Thrombospondin I fragment.
 XX
 DE
 XX
 KW Anti-angiogenic; carrier:DNA complex; tumour; gene therapy; human;
 KW thrombospondin I; melanoma; lung cancer; colon cancer;
 KW brain cancer; breast cancer.
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 160 /note= "encoded by CGG"
 FT Misc-difference 161 /note= "encoded by CTC"
 FT Misc-difference 162 /note= "encoded by TGC"
 FT Misc-difference 163 /note= "encoded by AAC"
 FT Misc-difference 177 /note= "encoded by GAA"

FT Misc-difference 178
 FT /note= "encoded by GCG"
 FT Misc-difference 179
 FT /note= "encoded by GCG"

PN EP921193-A1.

XX 09-JUN-1999.

XX 07-JAN-1998; 98EP-0100135.

XX 05-DEC-1997; 97US-0985526.

XX (MIXS/) MIXSON A J.

XX Mixson AJ;

XX WPI; 1999-315406/27.

XX N-PSDB; AAX58725.

XX Inhibition of growth of solid tumors

XX Disclosure; Page 17-18; 46pp; English.

XX This sequence represents an anti-angiogenic fragment of the
 CC thrombospondin I protein. The invention provides a carrier:DNA
 CC complex that comprises DNA (see AAX58725-42) coding for an
 CC anti-angiogenic protein or peptide, the complex being deliverable
 CC to the site of a tumour in vivo, and which additionally comprises
 CC regulatory elements for expressing the anti-angiogenic DNA in a
 CC tumour or tumour vasculature. The complex may also include DNA
 CC encoding a tumour suppressor protein, especially p53. The carrier
 CC is a liposome, cationic polymer, micelle, microsphere, virus, viral
 CC component, or a combination of these, and administration is by
 CC intravenous or intratumoral injection. The complexes are useful in
 CC gene therapy for inhibition of tumour growth. The types of tumors
 CC which may be treated include solid tumors such as melanomas and
 CC tumors in the lung, colon, brain and breast.

XX Sequence 218 AA;

Query Match 94.3%; Score 50; DB 20; Length 218;
 Best Local Similarity 85.7%; Pred. No. 1.9;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7

|||||

DB 82 WSPWSEW 88

RESULT 14

AAR40823
 ID AAR40823 standard; protein; 239 AA.

XX AAR40823;

XX 03-MAR-1994 (first entry)

XX Human thrombospondin 1.

XX hTSP-1; platelet glycoprotein; angiogenesis; neovascularisation;
 KW inhibitor; solid tumour; skin cancer; angiogenic dysfunction;
 KW melanoma; diabetic retinopathy; psoriasis; neovascular glaucoma;
 KW Kaposi's sarcoma; inflammation; retrolental fibroplasia.

XX Homo sapiens.

XX A09316716-A.

XX 02-SEP-1993.

XX 22-FEB-1993; 93WO-US01652.

XX 24-FEB-1992; 92US-0841656.
 XX (NOUN) UNIV NORTHWESTERN.
 PA Bouck NP, Frazier WA, Good DJ, Polverini PJ;
 XX WPI; 1993-288118/36.
 XX Compsn. for inhibiting angiogenesis - contg. a vascularisation
 PT inhibitor comprising a peptide capable of inhibiting
 PT vascularisation

XX Claim 10; Page 41-42; 51pp; English.

XX Certain fragments of thrombospondin, a glycoprotein found in the
 CC alpha granules of platelets, can inhibit vascularisation. Peptides
 CC derived from the hTSP sequence are useful for inhibiting
 CC neovascularisation, esp. in solid tumours such as melanomas. The
 CC inhibitory peptides can also be used in other diseases involving
 CC angiogenic dysfunction. See AAR40824-R40830.

XX Sequence 239 AA;

Query Match 94.3%; Score 50; DB 14; Length 239;
 Best Local Similarity 85.7%; Pred. No. 2.1;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSPWAEW 7

|||||

DB 71 WSPWSEW 77

RESULT 15

AAW40288

ID AAW40288 standard; Protein; 441 AA.

XX AAW40288;

XX 18-AUG-1998 (first entry)

XX Human concatamerised TSP1 protein.

XX TSP1; thrombospondin; anti-angiogenic; cationic vehicle; gene therapy;
 KW liposome; DNA complex; tumour suppressor protein; treatment; neoplastic;
 KW metabolic disease; concatamer; tumour.

XX Homo sapiens.

XX Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 160 /note= "encoded by GCG"

FT Misc-difference 221 /note= "encoded by GAC"

FT Misc-difference 383 /note= "encoded by GCG"

FT Region 223..227 /label= intervening sequence

XX EP819758-A2.

XX 21-JAN-1998.

XX 16-JUL-1997; 97EP-0112154.

XX 16-JUL-1996; 96US-0680845.

XX (MIXS/) MIXSON A J.

XX Mixson AJ;

DR WPI; 1998-078839/08.
DR N-PSDB; AAV10494.
XX
PT Complexes of DNA encoding anti-angiogenic peptide - with cationic
PT liposome(s) or cationic polymer, useful for, e.g. gene therapy of
PT tumours
XX
PS Claim 9; Page 6-7; 47pp; English.
XX
CC This protein sequence represents a concatamer of thrombospondin TSP1
CC which is used in a method to produce a cationic vehicle consisting
CC of a cationic liposome:DNA complex where the DNA encodes an
CC anti-angiogenic peptide or tumour suppressor protein. Such complexes are
CC used for treatment of neoplastic and metabolic diseases especially for
CC gene therapy of tumours.
XX
SQ Sequence 441 AA;

Query Match 94.3%; Score 50; DB 19; Length 441;
Best Local Similarity 85.7%; Pred. No. 4;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WSPWAEW 7
| | | | |
Db 82 wspwsew 88

Search completed: January 31, 2002, 09:23:07
Job time: 197 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:36:48 ; Search time 28.31 Seconds
(without alignments)
23.312 Million cell updates/sec

Title: US-09-536-087-7

Perfect score: 110

Sequence: 1 PATCANPSFVEGECPCSC 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	110	100.0	1172	1 TSP2_MOUSE	Q03350 mus musculus
2	105	95.5	1172	1 TSP2_HUMAN	P35442 homo sapien
3	98	89.1	1170	1 TSP2_BOVIN	Q95116 bos taurus
4	90	81.8	1178	1 TSP2_CHICK	P35440 gallus gall
5	63.5	57.7	810	1 NEL1_HUMAN	Q92832 homo sapien
6	59	53.6	1170	1 TSP1_BOVIN	Q28178 bos taurus
7	59	53.6	1170	1 TSP1_MOUSE	P07996 homo sapien
8	59	53.6	1170	1 TSP1_MOUSE	P35441 mus musculus
9	59	53.6	1173	1 TSP1_XENLA	P12105 gallus lae
10	59	53.6	1262	1 CAL3_CHICK	P12105 gallus gall
11	59	53.6	1459	1 CAL2_MOUSE	P28481 mus musculus
12	57	51.8	1496	1 CA25_HUMAN	P05997 homo sapien
13	56	50.9	1464	1 CAL1_MOUSE	P08121 mus musculus
14	55	50.0	1038	1 SOG_DROME	Q24025 drosophila
15	54	49.1	810	1 NEL1_RAT	Q62919 rattus norv
16	53	48.2	171	1 KR2A_SHEEP	P02438 ovis aries
17	53	48.2	941	1 CHR2_XENLA	Q91713 xenopus lae
18	52.5	47.7	816	1 NEL2_RAT	Q62918 rattus norv
19	52	47.3	1464	1 CAL1_HUMAN	P02452 homo sapien
20	51	46.4	156	1 KR2B_SHEEP	P02439 ovis aries
21	51	46.4	181	1 KR2D_SHEEP	P08131 ovis aries
22	51	46.4	1453	1 CAL1_CHICK	P02457 gallus gall
23	51	46.4	1466	1 CAL3_HUMAN	P02461 homo sapien
24	50	45.5	816	1 NEL2_HUMAN	Q99435 homo sapien
25	50	45.5	816	1 NEL2_MOUSE	Q61220 mus musculus
26	49.5	45.0	1250	1 TP3A_DROME	Q9ng98 drosophila
27	49	44.5	3951	1 VGF1_IVVB	P27920 avian infec
28	48.5	44.1	1056	1 MUC5_HUMAN	P98088 homo sapien
29	48.5	44.1	2700	1 ZAN_HUMAN	Q9y493 homo sapien
30	48	43.6	151	1 KR2C_SHEEP	P02440 ovis aries
31	48	43.6	1453	1 CAL1_MOUSE	P11087 mus musculus
32	47.5	43.2	194	1 CSR3_MOUSE	P50462 mus musculus
33	47.5	43.2	194	1 CSR3_RAT	P50463 rattus norv

RESULT 1

ID	TSP2_MOUSE	STANDARD;	PRT;	1172 AA.
AC	Q03350:			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	THROMBOSPONDIN 2 PRECURSOR.			
GN	THBS2 OR TSP2.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92147683; PubMed=1371115;			
RA	Laberty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,			
RA	Dixit V.M.;			
RT	"Characterization of mouse thrombospondin 2 sequence and expression			
RT	during cell growth and development."			
RL	J. Biol. Chem. 267:3274-3281(1992).			
RN	[2]			
RP	SEQUENCE OF 1-873 FROM N.A.			
RX	MEDLINE=91302287; PubMed=1712771;			
RA	Bornstein P., O'Rourke K., Wikstrom K., Wolf F.W., Katz R., Li P.,			
RA	Dixit V.M.;			
RT	"A second, expressed thrombospondin gene (Thbs2) exists in the mouse			
RT	genome."			
RL	J. Biol. Chem. 266:12821-12824(1991).			
CC	-1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND			
CC	CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,			
CC	LAMININ AND TYPE V COLLAGEN.			
CC	-1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.			
CC	-1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.			
CC	-1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.			
CC	-1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; L07803; AAA53064.1; -			
DR	EMBL; M64866; AAA40432.1; -			
DR	PIR; A42587; A42587.			
DR	PIR; A39851; A39851.			
DR	HSPP; P00740; IIXA.			
DR	MGI; MGI:98738; Thbs2.			
DR	InterPro; IPR000561; EGF-like.			
DR	InterPro; IPR001881; EGF_Ca.			
DR	InterPro; IPR001791; Laminin_G.			

ALIGNMENTS

```
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; VWFC.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 9.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00092; TSP1; 3.
DR PROSITE; PS01208; VWFC; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1172
FT DOMAIN 19 232
FT DOMAIN 318 375
FT DOMAIN 381 432
FT DOMAIN 437 493
FT DOMAIN 494 548
FT DOMAIN 549 589
FT DOMAIN 590 647
FT DOMAIN 648 692
FT DOMAIN 725 760
FT DOMAIN 761 783
FT DOMAIN 784 819
FT DOMAIN 820 842
FT DOMAIN 843 880
FT DOMAIN 881 916
FT DOMAIN 917 952
FT DOMAIN 953 1172
FT SITE 930
FT DISULFID 266 266
FT DISULFID 270 270
FT DISULFID 553 564
FT DISULFID 558 574
FT DISULFID 577 588
FT DISULFID 594 610
FT DISULFID 601 619
FT DISULFID 622 646
FT DISULFID 652 665
FT DISULFID 659 678
FT DISULFID 680 691
FT CARBOHYD 151 151
FT CARBOHYD 316 316
FT CARBOHYD 330 330
FT CARBOHYD 457 457
FT CARBOHYD 584 584
FT CARBOHYD 710 710
FT CARBOHYD 1069 1069
SQ SEQUENCE 1172 AA; 129911 MW; 7CE8E4E859822AB CRC64;

Query Match 100.08; Score 110; DB 1; Length 1172;
Best Local Similarity 100.08; Pred. No. 6.3e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PATCANPSEVEGCCPSC 18
Db 357 PATCANPSEVEGCCPSC 374
|||||
|||||

RESULT 2
ID TSP2_HUMAN STANDARD; PRT; 1172 AA.
AC P35442.
DT 01-JUN-1994 (Rel. 29, Created)
```

01-JUN-1994 (Rel. 29, Last sequence update)
20-AUG-2001 (Rel. 40, Last annotation update)
THROMBOSPONDIN 2 PRECURSOR.
THBS2 OR TSP2.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=94010892; PubMed=8406456;
Labell T.L., Byers P.H.;
"Sequence and characterization of the complete human thrombospondin 2
cDNA: potential regulatory role for the 3' untranslated region.";
Genomics 17:225-229(1993).
[2]
SEQUENCE OF 560-1172 FROM N.A.
TISSUE=Fibroblast;
RA Labell T.L., McGookey Milewicz D.J., Distèche C.M., Byers P.H.;
"Thrombospondin II: partial cDNA sequence, chromosome location, and
expression of a second member of the thrombospondin gene family in
humans.";
Genomics 12:421-429(1992).
CC FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
LAMININ AND TYPE V COLLAGEN.
CC SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC SIMILARITY: BELONGS TO THE THROMOSPONDIN FAMILY.
CC SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; L12350; AAA03703.1; -
EMBL; M81339; -; NOT_ANNOTATED_CDS.
PIR; A42173; A42173.
HSP; P00740; 11XA.
MIM; 188061; -
InterPro; IPR000561; EGF-like.
InterPro; IPR001881; EGF_Ca.
InterPro; IPR001791; Laminin_G.
InterPro; IPR000884; TSP1.
InterPro; IPR003129; TSPN.
InterPro; IPR001007; VWFC.
InterPro; IPR003367; tsp_3.
Pfam; PF00008; EGF; 2.
Pfam; PF02210; TSPN; 1.
Pfam; PF00090; tsp_1; 3.
Pfam; PF02412; tsp_3; 9.
Pfam; PF00093; vwc; 1.
SMART; SM00181; EGF; 3.
SMART; SM00209; TSP1; 3.
SMART; SM00210; TSPN; 1.
SMART; SM00214; VWC; 1.
PROSITE; PS00022; EGF_1; FALSE_NEG.
PROSITE; PS01186; EGF_2; 1.
PROSITE; PS00092; TSP1; 3.
PROSITE; PS01208; VWFC; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1172
FT DOMAIN 19 232
FT DOMAIN 318 375
FT DOMAIN 381 432
FT DOMAIN 437 493
FT DOMAIN 494 548
FT DOMAIN 549 589
FT DOMAIN 590 647
FT DOMAIN 648 692
FT DOMAIN 725 760
FT DOMAIN 761 783
FT DOMAIN 784 819
FT DOMAIN 820 842
FT DOMAIN 843 880
FT DOMAIN 881 916
FT DOMAIN 917 952
FT DOMAIN 953 1172
FT SITE 930
FT DISULFID 266 266
FT DISULFID 270 270
FT DISULFID 553 564
FT DISULFID 558 574
FT DISULFID 577 588
FT DISULFID 594 610
FT DISULFID 601 619
FT DISULFID 622 646
FT DISULFID 652 665
FT DISULFID 659 678
FT DISULFID 680 691
FT CARBOHYD 151 151
FT CARBOHYD 316 316
FT CARBOHYD 330 330
FT CARBOHYD 457 457
FT CARBOHYD 584 584
FT CARBOHYD 710 710
FT CARBOHYD 1069 1069
SQ SEQUENCE 1172 AA; 129911 MW; 7CE8E4E859822AB CRC64;

```

FT DOMAIN 381 432 TSP TYPE-1 1.
FT DOMAIN 437 493 TSP TYPE-1 2.
FT DOMAIN 494 548 TSP TYPE-1 3.
FT DOMAIN 549 589 EGF-LIKE 1.
FT DOMAIN 590 647 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 648 692 EGF-LIKE 3.
FT DOMAIN 725 760 TSP TYPE-3 1.
FT DOMAIN 761 783 TSP TYPE-3 2.
FT DOMAIN 784 819 TSP TYPE-3 3.
FT DOMAIN 820 842 TSP TYPE-3 4.
FT DOMAIN 843 880 TSP TYPE-3 5.
FT DOMAIN 881 916 TSP TYPE-3 6.
FT DOMAIN 917 952 TSP TYPE-3 7.
FT DOMAIN 953 1172 C-TERMINAL.
FT SITE 928 930 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 266 266 INTERCHAIN (PROBABLE).
FT DISULFID 270 270 INTERCHAIN (PROBABLE).
FT DISULFID 553 564 BY SIMILARITY.
FT DISULFID 558 574 BY SIMILARITY.
FT DISULFID 577 588 BY SIMILARITY.
FT DISULFID 594 610 BY SIMILARITY.
FT DISULFID 601 619 BY SIMILARITY.
FT DISULFID 622 646 BY SIMILARITY.
FT DISULFID 652 665 BY SIMILARITY.
FT DISULFID 659 678 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT CARBOHYD 151 151 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 584 584 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1069 1069 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1172 AA; 129955 MW; 2AC7BB230E44C6F5 CRC64;

Query Match 95.5%; Score 105; DB 1; Length 1172;
Best Local Similarity 94.4%; Pred. No. 3.1e-07;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PATCANSFVEGECPCSPC 18
Db 357 PATCANSFVEGECPCSPC 374
||||:|||||
||||:|||||

RESULT 3
TSP2_BOVIN STANDARD; PRT; 1170 AA.
AC Q95116; Q28180;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 2 PRECURSOR (CORTICOTROPIN-INDUCED SECRETED PROTEIN)
DE (CISP).
GN THB2 OR TSP2 OR TSP-2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Danik M., Chinn A., Lefeuvre M., Keramidas M., Aquesse-Germon S.,
RA Penhoat A., Chen H., Mosher D., Chambaz E.M., Feige J.J.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-522 FROM N.A.
RX MEDLINE=9631130; PubMed=8698834;
RA Lefeuvre B., Pellerin S., Keramidas M., Danik M., Chambaz E.M.,
RA Feige J.J.;
RT "Opposite regulation of thrombospondin-1 and corticotropin-induced
RT secreted protein/thrombospondin-2 expression by adrenocorticotrophic
RT hormone in adrenocortical cells."

```

```

RL J. Cell. Physiol. 167:164-172(1996).
RN [3]
RN SEQUENCE OF 318-831 FROM N.A.
RC TISSUE-Aortic endothelium;
RA Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
RT "Cloning and sequencing of bovine thrombospondin stimulatory effect of
RT TGF-beta."
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 WMFC DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X96540; CAA65385.1; -
CC EMBL; X87620; CAA60952.1; -
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR003129; TSPN.
CC InterPro; IPR001007; WMFC.
CC InterPro; IPR003367; tsp_3.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF02210; TSPN; 1.
CC Pfam; PF00090; tsp_1; 3.
CC Pfam; PF02412; tsp_3; 9.
CC Pfam; PF00093; WVC; 1.
CC SMART; SM00181; EGF; 3.
CC SMART; SM00209; TSP1; 3.
CC SMART; SM00210; TSPN; 1.
CC SMART; SM00214; WVC; 1.
CC PROSITE; PS00022; EGF_1; FALSE_NEG.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS00092; TSP1; 3.
CC PROSITE; PS01208; WMFC; 1.
CC Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
CC EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1170 THROMBOSPONDIN 2.
FT DOMAIN 19 232 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 318 375 WMFC.
FT DOMAIN 379 430 TSP TYPE-1 1.
FT DOMAIN 435 491 TSP TYPE-1 2.
FT DOMAIN 492 546 TSP TYPE-1 3.
FT DOMAIN 547 587 EGF-LIKE 1.
FT DOMAIN 588 645 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 646 690 EGF-LIKE 3.
FT DOMAIN 723 758 TSP TYPE-3 1.
FT DOMAIN 759 781 TSP TYPE-3 2.
FT DOMAIN 782 817 TSP TYPE-3 3.
FT DOMAIN 818 840 TSP TYPE-3 4.
FT DOMAIN 841 878 TSP TYPE-3 5.
FT DOMAIN 879 914 TSP TYPE-3 6.
FT DOMAIN 915 950 TSP TYPE-3 7.
FT DOMAIN 951 1170 C-TERMINAL.
FT SITE 926 928 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 266 266 INTERCHAIN (PROBABLE).
FT DISULFID 270 270 INTERCHAIN (PROBABLE).
FT DISULFID 551 562 BY SIMILARITY.
FT DISULFID 556 572 BY SIMILARITY.

```

FT DISULFID 575 586 BY SIMILARITY.
 FT DISULFID 592 608 BY SIMILARITY.
 FT DISULFID 599 617 BY SIMILARITY.
 FT DISULFID 620 644 BY SIMILARITY.
 FT DISULFID 650 663 BY SIMILARITY.
 FT DISULFID 657 676 BY SIMILARITY.
 FT DISULFID 678 689 BY SIMILARITY.
 FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 435 435 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 582 582 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 936 936 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 535 535 A -> V (IN REF. 3).
 FT CONFLICT 748 748 S -> T (IN REF. 3).
 SQ SEQUENCE 1170 AA; 129862 MW; 9CF1FBF5B89A051 CRC64;

Query Match 89.1%; Score 98; DB 1; Length 1170;
 Best Local Similarity 88.9%; Pred. No. 2.8e-06;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PATCANPSFVEGCCPSC 18
 ||||| |||||
 Db 357 PATCADPWFVEGCCPSC 374

RESULT 4

TSP2_CHICK STANDARD; PRT: 1178 AA.
 AC P35440;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE THROMBOSPONDIN 2 PRECURSOR.
 GN THB2 OR TSP2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91217026; PubMed-2022631;
 RA Lawler J., Duquette M., Ferro P.;
 RA "Cloning and sequencing of chicken thrombospondin.";
 RL J. Biol. Chem. 266:8039-8043(1991).
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 LAMININ AND TYPE V COLLAGEN.
 CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-1 DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL; M60853; AAA51437.1; .
 DR PIR; A39804; A39804.
 DR HSP; P00740; IIXA.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001791; Laminin_G.

DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR001007; VWFC.
 DR InterPro; IPR003367; tsp_3.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF00090; tsp_1; 3.
 DR Pfam; PF02412; tsp_3; 8.
 DR Pfam; PF00093; vwc; 1.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00001; EGF_like; 1.
 DR SMART; SM00209; TSP1; 3.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50092; TSP1; 3.
 DR PROSITE; PS01208; VWFC; 1.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 ?
 FT CHAIN ? 1178 THROMBOSPONDIN 2.
 FT DOMAIN ? 232 HEPARIN-BINDING (POTENTIAL).
 FT DOMAIN 324 381 VWFC.
 FT DOMAIN 387 438 TSP TYPE-1 1.
 FT DOMAIN 441 499 TSP TYPE-1 2.
 FT DOMAIN 500 553 TSP TYPE-1 3.
 FT DOMAIN 555 595 EGF-LIKE 1.
 FT DOMAIN 596 653 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 654 698 EGF-LIKE 3.
 FT DOMAIN 731 766 TSP TYPE-3 1.
 FT DOMAIN 767 789 TSP TYPE-3 2.
 FT DOMAIN 790 825 TSP TYPE-3 3.
 FT DOMAIN 826 848 TSP TYPE-3 4.
 FT DOMAIN 849 886 TSP TYPE-3 5.
 FT DOMAIN 887 922 TSP TYPE-3 6.
 FT DOMAIN 923 958 TSP TYPE-3 7.
 FT DOMAIN 959 1178 C-TERMINAL.
 FT SITE 934 935 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 559 570 BY SIMILARITY.
 FT DISULFID 564 580 BY SIMILARITY.
 FT DISULFID 583 594 BY SIMILARITY.
 FT DISULFID 600 616 BY SIMILARITY.
 FT DISULFID 607 625 BY SIMILARITY.
 FT DISULFID 628 652 BY SIMILARITY.
 FT DISULFID 658 671 BY SIMILARITY.
 FT DISULFID 665 684 BY SIMILARITY.
 FT DISULFID 686 697 BY SIMILARITY.
 FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 590 590 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 716 716 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1075 1075 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1178 AA; 131816 MW; F37E02F42C8717A2 CRC64;

Query Match 81.8%; Score 90; DB 1; Length 1178;
 Best Local Similarity 72.2%; Pred. No. 3.6e-05;
 Matches 13; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PATCANPSFVEGCCPSC 18
 |:::|||||
 Db 363 PVSCADPWFVEGCCPVC 380

RESULT 5
 ID NELL_HUMAN STANDARD; PRT: 810 AA.
 AC Q92832; Q9Y472;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)

20-AUG-2001 (Rel. 40, Last annotation update)
 PROTEIN KINASE C-BINDING PROTEIN NELL1 PRECURSOR (NEL-LIKE PROTEIN 1)
 (NEL-RELATED PROTEIN 1).
 GN NELL1 OR NR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97131504; PubMed=8975702;
 RA Watanabe T.K., Katagiri T., Suzuki M., Shimizu F., Fujiwara T.,
 RA Nakamoto N., Nakamura Y., Hirai Y., Maekawa H., Takahashi E.,
 RT "Cloning and characterization of two novel human cDNAs (NELL1 and
 RT NELL2) encoding proteins with six EGF-like repeats.";
 RL Genomics 38:273-276(1996).
 RN [2]
 RP SEQUENCE OF 383-810 FROM N.A.
 RA Ting K., Vastardis H., Mulliken J.B., Bertolami C., Wen Z.,
 RA Young M., Tieu A., Kwong E.;
 RT "Nell homolog gene expression in craniofacial anomalies.";
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: HOMOTRIMER. BINDS TO PKC BETA-1 (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -1- DISEASE: EXPRESSED IN CRANIOFACIAL ANOMALIES.
 CC -1- SIMILARITY: CONTAINS 1 TSP N-TERMINAL DOMAIN.
 CC -1- SIMILARITY: CONTAINS 5 VWFC DOMAINS.
 CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS
 IN POSITIONS 427 AND 771.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D83017; BAAL1680.1;
 DR EMBL: U57523; AAB06946.1; ALT_FRAME.
 DR HSP: P07204; ZADX.
 DR MIM: 602319;
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001791; Laminin_G.
 DR InterPro: IPR003129; TSPN.
 DR InterPro: IPR001007; VWFC.
 DR Pfam: PF00008; EGF; 4.
 DR Pfam: PF00093; vwc; 2.
 DR SMART: SM00179; EGF_Ca; 2.
 DR SMART: SM00001; EGF_like; 4.
 DR SMART: SM00282; Lamg; 1.
 DR SMART: SM00210; TSPN; 1.
 DR SMART: SM00214; vwc; 3.
 DR SMART: SM00011; vwc_def; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 3.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS01186; EGF_2; 3.
 DR PROSITE: PS01187; EGF_CA; 3.
 DR PROSITE: PS01208; VWFC; 2.
 KW Glycoprotein; EGF-like domain; Repeat; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 810
 FT PROTEIN KINASE C-BINDING PROTEIN NELL1.
 FT DOMAIN 81 230
 FT TSP N-TERMINAL.
 FT DOMAIN 273 331
 FT VWFC 1.
 FT DOMAIN 335 390
 FT VWFC 2.
 FT DOMAIN 391 433
 FT EGF-LIKE 1.
 FT DOMAIN 434 475
 FT EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 476 516
 FT EGF-LIKE 3.
 FT DOMAIN 515 547
 FT EGF-LIKE 4.

DOMAIN 549 595 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 DOMAIN 596 631 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 DOMAIN 632 687 VWFC 3.
 DOMAIN 687 750 VWFC 4.
 DOMAIN 752 807 VWFC 5.
 DISULFID 395 407 BY SIMILARITY.
 DISULFID 401 416 BY SIMILARITY.
 DISULFID 418 432 BY SIMILARITY.
 DISULFID 438 451 BY SIMILARITY.
 DISULFID 445 460 BY SIMILARITY.
 DISULFID 462 474 BY SIMILARITY.
 DISULFID 480 493 BY SIMILARITY.
 DISULFID 497 502 BY SIMILARITY.
 DISULFID 504 515 BY SIMILARITY.
 DISULFID 519 529 BY SIMILARITY.
 DISULFID 523 535 BY SIMILARITY.
 DISULFID 537 546 BY SIMILARITY.
 DISULFID 553 566 BY SIMILARITY.
 DISULFID 560 575 BY SIMILARITY.
 DISULFID 577 594 BY SIMILARITY.
 DISULFID 600 613 BY SIMILARITY.
 DISULFID 607 622 BY SIMILARITY.
 DISULFID 624 630 BY SIMILARITY.
 CARBOHYD 40 40 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 83 83 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 224 224 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 294 294 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 511 511 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 562 562 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 609 609 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 708 708 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 732 732 N-LINKED (GLCNAC. .) (POTENTIAL).
 CARBOHYD 758 758 N-LINKED (GLCNAC. .) (POTENTIAL).
 CONFLICT 383 383 N -> D (IN REF. 2).
 CONFLICT 573 573 Y -> H (IN REF. 2).
 CONFLICT 626 626 S -> C (IN REF. 2).
 SQ SEQUENCE 810 AA; 89606 MW; 549465EA3F7AEED0 CRC64;
 Query Match 57.7%; Score 63.5; DB 1; Length 810;
 Best Local Similarity 52.2%; Pred. No. 0.12;
 Matches 12; Conservative 1; Mismatches 5; Indels 5; Gaps 1;
 QY 1 PATCANPS-----FVEGECPCPC 18
 Db 727 PLTCPNLSCYEYTAIEGECPCPC 749
 RESULT 6
 TSP1_BOVIN STANDARD; PRT; 1170 AA.
 AC Q28178; Q28179;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE THROMBOSPONDIN 1 PRECURSOR.
 GN THBS1 OR TSP1 OR TSP-1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HOLSTEIN; TISSUE=Tooth;
 RX MEDLINE=98173773; PubMed=9507054;
 RA Ueno A., Yamashita K., Nagata T., Tsurumi C., Miwa Y., Kitamura S.,
 RA Inoue H.;
 RT "cDNA cloning of bovine thrombospondin 1 and its expression in
 RT odontoblasts and preodontoblasts.";
 RL Biochim. Biophys. Acta 1382:17-22(1998).

[2]
 RN SEQUENCE OF 1-18 AND 710-1170 FROM N.A.
 RP TISSUE-Aortic endothelium;
 RA Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 CC LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-
 CC V/BETA-3 AND ALPHA-IIB/BETA-3. MAY PLAY A ROLE IN DENTINOGENESIS
 CC AND/OR MAINTENANCE OF BENTIN AND DENTAL PULP.
 CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
 CC -1- TISSUE SPECIFICITY: ODONTOBLASTS.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 DR EMBL; AB005287; BAA21115.1; -;
 DR EMBL; X87618; CAA60950.1; -;
 DR EMBL; X87619; CAA60951.1; -;
 DR HSSP; P35555; LEMO.
 DR GlycoSuiteDB; Q281178; -;
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR001007; VWFC.
 DR InterPro; IPR003367; tsp_3.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF02210; tsp_1; 3.
 DR Pfam; PF02412; tsp_3; 16.
 DR Pfam; PF00093; vwc; 1.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00209; TSP1; 3.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; FALSE_NEG.
 DR PROSITE; PS01208; VWFC; 1.
 DR PROSITE; PS00092; TSP1; 3.
 DR Glycoprotein; Cell adhesion; Calcium-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1170
 FT THROMBOSPONDIN 1.
 FT HEPARIN-BINDING (POTENTIAL).
 FT VWFC.
 FT TSP TYPE-1 1.
 FT TSP TYPE-1 2.
 FT TSP TYPE-1 3.
 FT EGF-LIKE 1.
 FT EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT EGF-LIKE 3.
 FT EGF-LIKE 4.
 FT TSP TYPE-3 1.
 FT TSP TYPE-3 2.
 FT TSP TYPE-3 3.
 FT TSP TYPE-3 4.
 FT TSP TYPE-3 5.
 FT TSP TYPE-3 6.
 FT TSP TYPE-3 7.
 FT C-TERMINAL.
 FT CELL ATTACHMENT SITE (POTENTIAL).
 FT INTERCHAIN (PROBABLE).
 FT 270

FT DISULFID 274 INTERCHAIN (PROBABLE).
 FT DISULFID 551 BY SIMILARITY.
 FT DISULFID 572 BY SIMILARITY.
 FT DISULFID 575 BY SIMILARITY.
 FT DISULFID 586 BY SIMILARITY.
 FT DISULFID 592 BY SIMILARITY.
 FT DISULFID 599 BY SIMILARITY.
 FT DISULFID 620 BY SIMILARITY.
 FT DISULFID 644 BY SIMILARITY.
 FT DISULFID 657 BY SIMILARITY.
 FT DISULFID 678 BY SIMILARITY.
 FT CARBOHYD 248 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 360 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 708 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1067 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1085 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 805 S -> G (IN REF. 2).
 SQ SEQUENCE 1170 AA; 129533 MW; 0DD6ADF3E5FA031A CRC64;
 Query Match 53.6%; Score 59; DB 1; Length 1170;
 Best Local Similarity 53.3%; Pred. No. 0.66;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 Oy 4 CANPSFVGECCPSC 18
 I : I : |||||
 Db 358 CSNATVPDGECCPRC 372
 RESULT 7
 TSP1_HUMAN STANDARD; PRT; 1170 AA.
 ID TSP1_HUMAN AC P07996;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE THROMBOSPONDIN 1 PRECURSOR.
 GN THBS1 OR TSP1 OR TSP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Endothelial cells;
 RX MEDLINE=87057617; PubMed=2430973;
 RA Lawler J., Hynes R.O.;
 RT "The structure of human thrombospondin, an adhesive glycoprotein with
 RT multiple calcium-binding sites and homologies with several different
 RT proteins.";
 RL J. Cell Biol. 103:1635-1648(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89139590; PubMed=2918029;
 RA Hennessy S.W., Frazier B.A., Kim D.D., Deckwerth T.L.,
 RA Baumgartel D.M., Rotwein P., Frazier W.A.;
 RT "Complete thrombospondin mRNA sequence includes potential regulatory
 RT sites in the 3' untranslated region.";
 RL J. Cell Biol. 108:729-736(1989).
 RN [3]
 RP SEQUENCE OF 1-397 FROM N.A.
 RX MEDLINE=87157592; PubMed=3030396;
 RA Kobayashi S., Eden-McCutchan F., Framson P., Bornstein P.;
 RT "Partial amino acid sequence of human thrombospondin as determined by
 RT analysis of cDNA clones: homology to malarial circumsporozoite
 RT proteins.";
 RL Biochemistry 25:8418-8425(1986).
 RN [4]
 RP SEQUENCE OF 1-374 FROM N.A.
 RX MEDLINE=86287276; PubMed=3461443;
 RA Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
 RT "Characterization of a cDNA encoding the heparin and collagen binding
 RT domains of human thrombospondin.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).

[5]
 RN SEQUENCE OF 1-166 FROM N.A.
 RX MEDLINE-89291870; PubMed-2544587;
 RA Laherty C.D., Gierman T.M., Dixit V.M.;
 RT "Characterization of the promoter region of the human thrombospondin
 gene. DNA sequences within the first intron increase transcription.";
 RL J. Biol. Chem. 264:11222-11227(1989).
 [6]
 RN SEQUENCE OF 1028-1170 FROM N.A.
 RA la Fleur M., Jobin C., Gauthier J., Kreis C.G.;
 RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-
 V/BETA-3 AND ALPHA-11B/BETA-3.
 CC -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
 CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
 CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
 CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; M25631; AAA36741.1; -;
 DR EMBL; X04665; CAA28370.1; -;
 DR EMBL; X14787; CAA32889.1; -;
 DR EMBL; J04835; AAA61178.1; -;
 DR EMBL; M99425; AAB59366.1; -;
 DR PIR; A05172; A05172.
 DR PIR; A25812; A25812.
 DR PIR; A26155; A26155.
 DR PIR; A30140; A30140.
 DR PIR; A34274; A34274.
 DR HSSP; P35555; LEMO.
 DR MIM; 188060; -;
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR000884; TSPI.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR001007; VWFC.
 DR InterPro; IPR003367; tsp_3.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF00090; tsp_1; 3.
 DR Pfam; PF02412; tsp_3; 8.
 DR Pfam; PF00093; vwc; 1.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00209; TSP1; 3.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS00092; TSP1; 3.
 DR PROSITE; PS01208; VWFC; 1.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1170 THROMBOSPONDIN 1.
 FT DOMAIN 19 232 HEPARIN-BINDING (POTENTIAL).
 FT DOMAIN 316 373 VWFC.
 FT DOMAIN 379 430 TSP TYPE-1 1.
 FT DOMAIN 435 491 TSP TYPE-1 2.
 FT DOMAIN 492 548 TSP TYPE-1 3.
 FT DOMAIN 549 587 EGF-LIKE 1.
 FT DOMAIN 588 645 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 646 690 EGF-LIKE 3.
 FT DOMAIN 723 758 TSP TYPE-3 1.
 FT DOMAIN 759 781 TSP TYPE-3 2.
 FT DOMAIN 782 817 TSP TYPE-3 3.
 FT DOMAIN 818 840 TSP TYPE-3 4.
 FT DOMAIN 841 878 TSP TYPE-3 5.
 FT DOMAIN 879 914 TSP TYPE-3 6.
 FT DOMAIN 915 950 TSP TYPE-3 7.
 FT DOMAIN 951 1170 C-TERMINAL.
 FT SITE 926 928 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 270 270 INTERCHAIN (PROBABLE).
 FT DISULFID 274 274 INTERCHAIN (PROBABLE).
 FT DISULFID 551 562 BY SIMILARITY.
 FT DISULFID 556 572 BY SIMILARITY.
 FT DISULFID 575 586 BY SIMILARITY.
 FT DISULFID 592 608 BY SIMILARITY.
 FT DISULFID 599 617 BY SIMILARITY.
 FT DISULFID 620 644 BY SIMILARITY.
 FT DISULFID 650 663 BY SIMILARITY.
 FT DISULFID 657 676 BY SIMILARITY.
 FT DISULFID 678 689 BY SIMILARITY.
 FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 84 84 T -> A (IN REF. 2, 3 AND 4).
 FT CONFLICT 523 523 T -> A (IN REF. 2).
 SQ SEQUENCE 1170 AA; 129412 MW; 69B3DE5AE3A395E CRC64;
 Query Match 53.6%; Score 59; DB 1; Length 1170;
 Best Local Similarity 53.3%; Pred. No. 0.66;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 4 CANPSFVEGECPCSC 18
 Db 358 CSNATVPDGECCPRC 372
 RESULT 8
 TSP1_MOUSE STANDARD; PRT; 1170 AA.
 ID TSP1_MOUSE
 AC P35441;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE THROMBOSPONDIN 1 PRECURSOR.
 GN THBS1 OR TSP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92128941; PubMed=1774063;
 RA Lawler J., Duquette M., Ferro P., Copeland N.G., Gilbert D.J.,
 RA Jenkins N.A.;
 RT "Characterization of the murine thrombospondin gene.";
 RL Genomics 11:587-600(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92147683; PubMed=1371115;
 RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
 RA Dixit V.M.;
 RT "Characterization of mouse thrombospondin 2 sequence and expression
 during cell growth and development.";
 RL J. Biol. Chem. 267:3274-3281(1992).
 RN [3]
 RP SEQUENCE OF 1-490 FROM N.A.
 RX MEDLINE-90375546; PubMed=2398070;
 RA Bornstein P., Alfi D., Devarayalu S., Framson P., Li P.;
 RT "Characterization of the mouse thrombospondin gene and evaluation of
 the role of the first intron in human gene expression.";

FT	DOMAIN	19	232	HEPARIN-BINDING (POTENTIAL).
FT	DOMAIN	316	373	WFCC.
FT	DOMAIN	379	430	TSP TYPE-1 1.
FT	DOMAIN	435	491	TSP TYPE-1 2.
FT	DOMAIN	492	548	TSP TYPE-1 3.
FT	DOMAIN	549	587	EGF-LIKE 1.
FT	DOMAIN	588	645	EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	646	690	EGF-LIKE 3.
FT	DOMAIN	723	758	TSP TYPE-3 1.
FT	DOMAIN	759	781	TSP TYPE-3 2.
FT	DOMAIN	782	817	TSP TYPE-3 3.
FT	DOMAIN	841	840	TSP TYPE-3 4.
FT	DOMAIN	848	878	TSP TYPE-3 5.
FT	DOMAIN	879	914	TSP TYPE-3 6.
FT	DOMAIN	915	950	TSP TYPE-3 7.
FT	DOMAIN	951	1170	C-TERMINAL.
FT	FT SITE	926	928	CELL ATTACHMENT SITE (POTENTIAL).
FT	DISULFID	270	270	INTERCHAIN (PROBABLE).
FT	DISULFID	274	274	INTERCHAIN (PROBABLE).
FT	DISULFID	551	562	BY SIMILARITY.
FT	DISULFID	556	572	BY SIMILARITY.
FT	DISULFID	575	586	BY SIMILARITY.
FT	DISULFID	592	608	BY SIMILARITY.
FT	DISULFID	599	617	BY SIMILARITY.
FT	DISULFID	620	644	BY SIMILARITY.
FT	DISULFID	650	663	BY SIMILARITY.
FT	DISULFID	657	676	BY SIMILARITY.
FT	DISULFID	678	689	BY SIMILARITY.
FT	CARBOHYD	248	248	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	360	360	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	708	708	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1067	1067	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CONFLICT	1025	1025	F -> L (IN REF. 2).
SQ	SEQUENCE	1170 AA; 129646 MW; 0443E493615E7F06 CRC64;		

Query Match	53.6%;	Score 59;	DB 1;	Length 1170;
Best Local Similarity	53.3%;	Pred. No. 0.66;		
Matches	8;	Conservative	3;	Mismatches 4; Indels 0; Gaps

Qy	4	CANPSFVEGECPCSC	18
		: :	
Db	358	CSNATVPDGECCPRC	372

RESULT	9
TSPL_XENLA	
ID	TSPL_XENLA STANDARD; PRT; 1173 AA.
AC	P35448;
DT	01-JUN-1994 (Rel. 29, Created)
DT	01-JUN-1994 (Rel. 29, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	DE THROMBOSPONDIN 1 PRECURSOR.
GN	THBS1 OR TSPI.
OS	Xenopus laevis (African clawed frog).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC	Xenopodinae; Xenopus.
OC	NCBI_TaxID=8355;
RP	[1]
RP	SEQUENCE FROM N.A.
RA	Urby L.A., Ramos J., Duquette M., Desimone D.W., Lawler J.;
RA	Submitted (XXX-1993) to the EMBL/GenBank/DBJ databases.
RL	-!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC	CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN-
CC	LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-
CC	V/BETA-3 AND ALPHA-IIB/BETA-3 (BY SIMILARITY).
CC	-!- SUBUNIT: HOMOPHIMER; DISULFIDE-LINKED.
CC	-!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC	-!- SIMILARITY: CONTAINS 1 WFCC DOMAIN.
CC	-!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC	-!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC	-!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L04278; ; NOT_ANNOTATED_CDS.
HSSP; P00740; LIXA.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR003129; TSPN.
DR InterPro: IPR001007; VWFC.
DR InterPro: IPR003367; tsp_3.
DR Pfam: PF00008; EGF; 2.
DR Pfam: PF02210; TSPN; 1.
DR Pfam: PF00090; tsp_1; 3.
DR Pfam: PF02412; tsp_3; 8.
DR Pfam: PF00093; vwc; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00092; TSP1; 3.
DR PROSITE; PS01208; WFC; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1173
FT DOMAIN 23 235
FT DOMAIN 319 376
FT DOMAIN 382 433
FT DOMAIN 438 494
FT DOMAIN 495 546
FT DOMAIN 550 590
FT DOMAIN 591 648
FT DOMAIN 649 693
FT DOMAIN 726 761
FT DOMAIN 762 784
FT DOMAIN 785 820
FT DOMAIN 821 843
FT DOMAIN 844 881
FT DOMAIN 882 917
FT DOMAIN 918 953
FT DOMAIN 954 1173
FT SITE 929 931
FT DISULFID 554 565
FT DISULFID 559 575
FT DISULFID 578 589
FT DISULFID 595 611
FT DISULFID 602 620
FT DISULFID 623 647
FT DISULFID 653 666
FT DISULFID 660 679
FT DISULFID 681 692
FT CARBOHYD 155 155
FT CARBOHYD 158 158
FT CARBOHYD 250 250
FT CARBOHYD 363 363
FT CARBOHYD 705 705
FT CARBOHYD 711 711
FT CARBOHYD 1070 1070
SQ SEQUENCE 1173 AA; 130019 MW; A9F036D6516C0F24 CRC64;

Query Match

53.6%; Score 59; DB 1; Length 1173;

Best Local Similarity 53.3%; Pred. No. 0.66;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 4 CANPSFVEGECPCSC 18
DB 361 CTNATPDGECPCRC 375
RESULT 10
CAL3_CHICK STANDARD; PRT; 1262 AA.
AC P12105; P79758; P79759; Q90794; Q92029;
DT 01-OCT-1989 (Rel. 12, Created)
DT 20-OCT-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 1(III) CHAIN PRECURSOR (FRAGMENTS).
GN COL3A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE OF 1-886 FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94266842; PubMed=8206952;
RA Nah H.-D., Niu Z., Adams S.L.;
RT "An alternative transcript of the chick type III collagen gene that
does not encode type III collagen.";
RL J. Biol. Chem. 269:16443-16448(1994).
RN [2]
RP SEQUENCE OF 29-96; 332-397; 431-484; 503-535 AND 869-976 FROM N.A.
RX MEDLINE=84270696; PubMed=6547770;
RA Yamada Y., Lian G., Mudryj M., Obici S., de Crombrughe B.;
RT "Conservation of the sizes for one but not another class of exons in
two chick collagen genes.";
RL Nature 310:333-337(1984).
RN [3]
RP SEQUENCE OF 977-1262 FROM N.A.
RX MEDLINE=83220816; PubMed=6856474;
RA Yamada Y., Kuhn K., de Crombrughe B.;
RT "A conserved nucleotide sequence, coding for a segment of the C-
propeptide, is found at the same location in different collagen
genes.";
RL Nucleic Acids Res. 11:2733-2744(1983).
CC -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC -1- ALONG WITH TYPE I COLLAGEN.
CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC ALSO CROSS-LINKED VIA HYDROXYLYSINES.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; U07973; AAA83407.1; -;
DR EMBL; X00822; CAB52686.1; -;
DR EMBL; X00823; CAB52686.1; JOINED.
DR EMBL; X00824; CAA25396.1; ALT SEQ.
DR EMBL; X00823; CAA25396.1; JOINED.
DR EMBL; X00822; CAA25396.1; JOINED.
DR EMBL; X00826; CAA25397.1; ALT SEQ.
DR EMBL; X00825; CAA25397.1; JOINED.
DR EMBL; X00827; CAA25398.1; -;
DR EMBL; X00828; CAA25399.1; -;
DR EMBL; X00830; CAA25401.1; -;

```

DR EMBL; X00831; CAA25402.1; -.
DR EMBL; K02302; AAD15299.1; -.
DR EMBL; K02301; AAD15298.1; -.
DR EMBL; M36662; ARA18519.1; ALT_SEQ.
DR PIR; A05269; A05269.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR Pfam; PF00093; vwc; 1.
DR Pfam; PF01391; Collagen; 11.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWC; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 23
FT PROPEP 24 144
FT CHAIN 145 1003
FT PROPEP 1004 1262
FT DOMAIN 29 88
FT DOMAIN 145 164
FT DOMAIN 165 994
FT DOMAIN 995 1003
FT NON_CONS 886 887
FT DISULFID 922 923
FT DISULFID 994 994
FT DISULFID 995 995
FT MOD_RES 262 262
FT MOD_RES 283 283
FT MOD_RES 859 859
FT CARBOHYD 1163 1163
FT CONFLICT 96 96
FT CONFLICT 1132 1132
SQ SEQUENCE 1262 AA; 121249 MW; 96ABE7B2E9DEB43D CRC64;

Query Match 53.6%; Score 59; DB 1; Length 1262;
Best Local Similarity 60.0%; Pred. No. 0.71;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 CANPSFVEGECPC 18
DB 73 CPNPEIPGECPCV 87

RESULT 11
CA12_MOUSE STANDARD; PRT; 1459 AA.
AC P28481;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 1(II) CHAIN PRECURSOR [CONTAINS: CHONDROCALCIN].
GN COL2A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RX SEQUENCE FROM N.A.
MEDLINE=91358489; PubMed=1885613;
RA Metsaranta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Mouse type II collagen gene. Complete nucleotide sequence, exon
structure, and alternative splicing."
RL J. Biol. Chem. 266:16862-16869(1991).
RN [2]
RX SEQUENCE OF 1455-1459 FROM N.A.

MEDLINE=91274355; PubMed=2054384;
RA Metsaranta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Specific hybridization probes for mouse type I, II, III and IX
collagen mRNAs."
RL Biochim. Biophys. Acta 1089:241-243(1991).
CC -!- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.
CC -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- SIMILARITY: CONTAINS 1 VWC DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M65161; AAA68100.1; -.
CC EMBL; X57982; CAA41047.1; -.
CC MGD; MGI:88452; Col2a1.
CC InterPro; IPR000087; Collagen.
CC InterPro; IPR000885; Fib_collagen_C.
CC Pfam; PF01410; COLFI; 1.
CC Pfam; PF01391; Collagen; 17.
CC Pfam; PF00093; vwc; 1.
CC ProDom; PD002078; Fib_collagen_C; 1.
CC SMART; SM00038; COLFI; 1.
CC SMART; SM00214; VWC; 1.
CC PROSITE; PS01208; VWC; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Cartilage; Signal; Alternative splicing.
FT SIGNAL 1 25
FT PROPEP 26 153
FT CHAIN 154 1213
FT PROPEP 1214 1459
FT DOMAIN 32 89
FT DOMAIN 173 1186
FT DOMAIN 1187 1213
FT VARSPLIC 29 29
FT VARSPLIC 30 98
SQ SEQUENCE 1459 AA; 139154 MW; F6C84FA7C532E7F2 CRC64;

Query Match 53.6%; Score 59; DB 1; Length 1459;
Best Local Similarity 60.0%; Pred. No. 0.79;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 CANPSFVEGECPC 18
DB 74 CLNPEIPGECPC 88

RESULT 12
CA25_HUMAN STANDARD; PRT; 1496 AA.
ID CA25_HUMAN
AC P05997;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 2(V) CHAIN PRECURSOR.
GN COL5A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX SEQUENCE OF 1-463 FROM N.A.
MEDLINE=89123368; PubMed=2914927;

```

RA Woodbury D., Benson-Chanda V., Ramirez F.;
 RT "Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms
 to the structural criteria of a fibrillar procollagen molecule.";
 RL J. Biol. Chem. 264:2735-2738(1989).
 RN [2]
 RP SEQUENCE OF 398-1496 FROM N.A.
 RA MEDLINE=87146331; PubMed=3029669;
 RX Well D., Bernard M.P., Gargano S., Ramirez F.;
 RT "The pro alpha 2(V) collagen gene is evolutionarily related to the
 major fibrillar-forming collagens.";
 RL Nucleic Acids Res. 15:181-198(1987).
 RN [3]
 RP SEQUENCE OF 1227-1496 FROM N.A.
 RA MEDLINE=85289337; PubMed=2411731;
 RX Myers J.C., Loidl H.R., Seyer J.M., Dion A.S.;
 RT "Complete primary structure of the human alpha 2 type V procollagen
 COOH-terminal propeptide.";
 RL J. Biol. Chem. 260:11216-11222(1985).
 RN [4]
 RP SEQUENCE OF 1449-1496 FROM N.A.
 RA MEDLINE=89138450; PubMed=3224983;
 RX Tsipouras P., Schwartz R.C., Liddell A.C., Salkeld C.S., Weil D.,
 Ramirez F.;
 RT "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,
 located on the long arm of human chromosome 2.";
 RL Genomics 3:275-277(1988).
 RN [5]
 RP SEQUENCE OF 208-227.
 RA TISSUE=Placenta;
 RX MEDLINE=92239022; PubMed=1571108;
 Mann K.;
 RT "Isolation of the alpha 3-chain of human type V collagen and
 characterization by partial sequencing.";
 RL Biol. Chem. Hoppe-Seyler 373:69-75(1992).
 RN [6]
 RP SEQUENCE OF 288-297 AND 606-617.
 RA TISSUE=Bone;
 RX MEDLINE=94237164; PubMed=8181482;
 RA Moradi-Ameli M., Rousseau J.C., Klemm J.P., Champilaud M.F.,
 Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;
 RT "Diversity in the processing events at the N-terminus of type-V
 collagen.";
 RL Eur. J. Biochem. 221:987-995(1994).
 CC -1- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 (FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE
 COMPONENT OF NEARLY UBIQUITOUS DISTRIBUTION. TYPE V COLLAGEN BINDS
 TO DNA, HEPARAN SULFATE, THROMBOSPONDIN, HEPARIN, AND INSULIN.
 CC -1- SUBUNIT: TRIMERS OF TWO ALPHA 1(V) AND ONE ALPHA 2(V) CHAINS IN
 MOST TISSUES AND TRIMERS OF ONE ALPHA 1(V), ONE ALPHA 2(V), AND
 ONE ALPHA 3(V) CHAINS IN PLACENTA.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J04478; AAA51859.1; -;
 DR ENBL: X04758; CAA28454.1; -;
 DR EMBL: M11718; AAA52058.1; -;
 DR PIR: A25374; A25374.
 DR PIR: A25874; A25874.
 DR PIR: A30017; A30017.
 DR PIR: A31427; A31427.
 DR MIM: 120190; -;
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR000885; Fib.collagen_C.
 DR InterPro: IPR001007; VWFC.

DR Pfam: PF01410; COLFI; 1.
 DR Pfam: PF01391; Collagen; 18.
 DR Pfam: PF00093; vwc; 1.
 DR ProDom: PD002078; Fib_collagen_C; 1.
 DR SMART: SM00038; COLFI; 1.
 DR SMART: SM00214; VWC; 1.
 DR PROSITE: PS01208; VWFC; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 26
 FT CHAIN 27 1226 COLLAGEN ALPHA 2(V) CHAIN.
 FT PROPEP 1227 1496 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 39 97 VWFC.
 FT MOD_RES 290 290 HYDROXYLATION.
 FT MOD_RES 293 293 HYDROXYLATION.
 FT MOD_RES 296 296 HYDROXYLATION.
 FT MOD_RES 608 608 HYDROXYLATION.
 FT MOD_RES 614 614 HYDROXYLATION.
 FT CONFLICT 292 292 A -> P (IN REF. 6).
 FT CONFLICT 1418 1418 K -> T (IN REF. 3).
 FT CONFLICT 1438 1438 F -> S (IN REF. 3).
 FT CONFLICT 1460 1460 E -> Q (IN REF. 4).
 FT CONFLICT 1496 1496 V -> A (IN REF. 4).
 SQ SEQUENCE 1496 AA; 144720 MW; 82827C17A8644F5A CRC64;
 Query Match 51.88; Score 57; DB 1; Length 1496;
 Best Local Similarity 60.08; Pred. No. 1.5;
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 QY 4 CANPSFVEGECPC 18
 ||| |||||
 DB 82 CADPVTTPGECPCVC 96
 RESULT 13
 ID CA13_MOUSE STANDARD; PRT; 1464 AA.
 AC P08121; O61429; Q9CRN7;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 1(III) CHAIN PRECURSOR.
 GN COL3A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL X DBA; TISSUE=Embryo;
 RX MEDLINE=95011609; PubMed=7926795;
 RA Toman D., de Crombrughe B.;
 RT "The mouse type-III procollagen-encoding gene: genomic cloning and
 complete DNA sequence.";
 RL Gene 147:161-168(1994).
 RN [2]
 RP SEQUENCE OF 1-488 FROM N.A.
 RX MEDLINE=88167858; PubMed=3443309;
 RA Wood L., Theriault N., Vogel G.;
 RT "Complete nucleotide sequence of the N-terminal domains of the murine
 alpha-1 type-III collagen chain.";
 RL Gene 61:225-230(1987).
 RN [3]
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=85131189; PubMed=3972847;
 RA Liao G., Mudryj M., de Crombrughe B.;
 RT "Identification of the promoter and first exon of the mouse alpha 1
 (III) collagen gene.";
 RL J. Biol. Chem. 260:3773-3777(1985).
 RN [4]
 RP SEQUENCE OF 810-1464 FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic head;

RX MEDLINE-21085660; PubMed-11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Glssi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita C., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,
 RA "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 RN [5]
 RP SEQUENCE OF 1442-1464 FROM N.A.
 RC STRAIN-C57BL;
 RX MEDLINE-91274355; PubMed-2054384;
 RA Metsaeranta M., Toman D., de Crombrughe B., Vuorio E.;
 RT "Specific hybridization probes for mouse type I, II, III and IX
 RT collagen mRNAs";
 RL Biochim. Biophys. Acta 1089:241-243(1991).
 CC -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
 CC -!- ALONG WITH TYPE I COLLAGEN.
 CC -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
 CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
 CC ALSO CROSS-LINKED VIA HYDROXYLYSINES.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC O-LINKED GLYCAN CONSISTS OF GLC-GAL DISACCHARIDE (BY SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 1 WFPC DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X52046; CAA36279.1; -;
 DR EMBL; M18933; AAA37338.1; -;
 DR EMBL; K03037; -; NOT_ANNOTATED_CDS.
 DR EMBL; AK019448; BAB31724.1; -;
 DR EMBL; X57983; CAA41048.1; -;
 DR PIR; A22287; A22287.
 DR PIR; A27353; A27353.
 DR PIR; S16373; S16373.
 DR MGB; MGI:88453; Col3a1.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR000885; Fib.collagen_C.
 DR InterPro; IPR001007; WFPC.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 17.
 DR ProDom; PD002078; Fib.collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; WVC; 1.
 DR PROSITE; PS01208; WFPC; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 23
 FT PROPEP 24 154 AMINO-TERMINAL PROPEPTIDE.
 FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.
 FT PROPEP 1204 1464 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 31 90 WFPC.
 FT DOMAIN 155 169 NONHELICAL REGION (N-TERMINAL).

FT DOMAIN 170 1195 TRIPLE-HELICAL REGION.
 FT DOMAIN 1196 1464 NONHELICAL REGION (C-TERMINAL).
 FT CARBOHYD O-LINKED (GAL. . .) (BY SIMILARITY).
 FT MOD_RES 262 262 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 262 262 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 283 283 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 859 859 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 976 976 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 1093 1093 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 1105 1105 HYDROXYLATION (BY SIMILARITY).
 FT DISULFID 1195 1195 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 1196 1196 INTERCHAIN (BY SIMILARITY).
 SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;

 Query Match 50.9%; Score 56; DB 1; Length 1464;
 Best Local Similarity 50.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

 QY 1 PATCANPSFVEGECPCSP 18
 Db 72 PLDCPNPEIPFGECCAIC 89

 RESULT 14
 SOG_DROME
 ID SOG_DROME STANDARD; PRT; 1038 AA.
 AC Q24025; Q9VXST;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DORSAL-VENTRAL PATTERNING PROTEIN SOG (SHORT GASTRULATION PROTEIN).
 GN SOG OR CG9224.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-95047398; PubMed-7958919;
 RA Francois V., Solloway M., O'Neill J.W., Emery J., Bier E.;
 RT "Dorsal-ventral patterning of the Drosophila embryo depends on a
 RT putative negative growth factor encoded by the short gastrulation
 RT gene";
 RL Genes Dev. 8:2602-2616(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.I., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Leal Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: PUTATIVE NEGATIVE GROWTH FACTOR; ANTAGONIST OF DPP, A
CC PROTEIN INVOLVED IN PATTERNING THE DORSAL REGION AND IN THE
CC DEVELOPMENT OF THE NEUROECTODERM.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC -!- TISSUE SPECIFICITY: ABUTS THE DORSAL DPP-EXPRESSING CELLS IN A
CC LATERAL STRIPE 14-16 CELLS WIDE. LATER IN EMBRYOGENESIS IT IS
CC EXPRESSED IN NEUROECTODERM AND IN THE ENDODERM SPACED ALONG THE
CC ANTERIOR-POSTERIOR AXIS OF THE DEVELOPING GUT.
CC -!- DEVELOPMENTAL STAGE: EMBRYOGENESIS.
CC -!- SIMILARITY: BELONGS TO THE CHORDIN FAMILY.
CC -!- SIMILARITY: CONTAINS 4 VFMC DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U18774; AAA89117.1; -;
DR EMBL; AF003499; AAF48481.1; -;
DR FlyBase; FBgn0003463; sog.
DR InterPro; IPR001007; VFMC.
DR Pfam; PF000093; vwc; 4.
DR SMART; SM00214; VWC; 3.
DR SMART; SM00011; VWC_def; 1.
DR PROSITE; PS01208; VWC; 2.
DR Transmembrane; Developmental protein; Repeat; Growth factor;
KW Growth regulation; Signal-anchor.
FT DOMAIN 1 53
FT TRANSMEM 54 74
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT CYTOPLASMIC (POTENTIAL).
FT EXTRACELLULAR (POTENTIAL).
FT VFMC 1.
FT SR1.
FT SR2.
FT SR3.
FT VFMC 2.
FT DOMAIN 742 804
FT DOMAIN 830 899
FT DOMAIN 939 1020
FT CARBOHYD 179 179
FT CARBOHYD 287 287
FT CARBOHYD 520 520
FT CARBOHYD 666 666
FT CARBOHYD 752 752
FT CARBOHYD 821 821
FT SEQUENCE 1038 AA; 115514 MW; B0E833AFD79A9037 CRC64;

Query Match 50.0%; Score 55; DB 1; Length 1038;
Best Local Similarity 44.4%; Pred. No. 2.1;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
QY 1 PATCANPSFVEGCCPSC 18
||| :|| :||| :|
DB 157 PAKCDPISLPGKCKTC 174

RESULT 15
NEL1 RAT
ID NEL1 RAT STANDARD; PRT; 810 AA.
AC Q62919;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN KINASE C-BINDING PROTEIN NEL1 PRECURSOR (NEL-LIKE PROTEIN 1).
GN NEL1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX MEDLINE=20017976; PubMed=10548494;
RA Kuroda S., Oyama M., Kawakami M., Kanayama N., Tanizawa K., Saito N.,
RA Abe T., Matsuhashi S., Ting K.;
RT "Biochemical characterization and expression analysis of neural
RT thrombospondin-1-like proteins NEL1 and NEL2.";
RL Biochem. Biophys. Res. Commun. 265:79-86(1999).
CC -!- SUBUNIT: HOMOTRIMER. BINDS TO PKC BETA-1.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: CONTAINS 1 TSP N-TERMINAL DOMAIN.
CC -!- SIMILARITY: CONTAINS 5 VFMC DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U48246; AAC72252.1; -;
DR HSSP; P07204; 2ADX.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR001993; Mitoch_carrier.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; VFMC.
DR Pfam; PF00008; EGF; 8.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00093; VWC; 5.
DR SMART; SM00179; EGF_Ca; 2.
DR SMART; SM00001; EGF_like; 4.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01187; EGF_Ca; 3.
DR PROSITE; PS01208; VWC; 2.
KW Glycoprotein; EGF-like domain; Repeat; Signal.
FT SIGNAL 1 16
FT CHAIN 17 810
FT DOMAIN 81 230
FT DOMAIN 273 331
FT DOMAIN 335 390
FT DOMAIN 391 433
FT DOMAIN 434 475
FT DOMAIN 476 516
FT DOMAIN 515 547
FT DOMAIN 549 595
FT DOMAIN 596 631
FT DOMAIN 632 687
FT DOMAIN 692 750
FT DOMAIN 752 807
FT PROTEIN KINASE C-BINDING PROTEIN NEL1.
FT TSP N-TERMINAL.
FT VFMC 1.
FT VFMC 2.
FT EGF-LIKE 1.
FT EGF-LIKE 2.
FT EGF-LIKE 3.
FT EGF-LIKE 4.
FT EGF-LIKE 5.
FT EGF-LIKE 6.
FT VFMC 3.
FT VFMC 4.
FT VFMC 5.

```
FT DISULFID 395 407 BY SIMILARITY.
FT DISULFID 401 416 BY SIMILARITY.
FT DISULFID 418 432 BY SIMILARITY.
FT DISULFID 438 451 BY SIMILARITY.
FT DISULFID 445 460 BY SIMILARITY.
FT DISULFID 462 474 BY SIMILARITY.
FT DISULFID 480 493 BY SIMILARITY.
FT DISULFID 487 502 BY SIMILARITY.
FT DISULFID 504 515 BY SIMILARITY.
FT DISULFID 519 529 BY SIMILARITY.
FT DISULFID 523 535 BY SIMILARITY.
FT DISULFID 537 546 BY SIMILARITY.
FT DISULFID 553 566 BY SIMILARITY.
FT DISULFID 560 575 BY SIMILARITY.
FT DISULFID 577 594 BY SIMILARITY.
FT DISULFID 600 613 BY SIMILARITY.
FT DISULFID 607 622 BY SIMILARITY.
FT DISULFID 624 630 BY SIMILARITY.
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 224 224 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 562 562 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 609 609 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 810 AA; 89212 MW; 46F09C466AF9AB0B CRC64;
```

Query Match 49.1%; Score 54; DB 1; Length 810;
Best Local Similarity 53.3%; Pred. No. 2.4;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 4 CANPSFVEGCCPSC 18
| : ||||| |
Db 735 CEYAMFEGECCPRC 749

Search completed: January 31, 2002, 09:36:49
Job time: 768 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:23:58 ; Search time 43.81 seconds
(without alignments)
3.082 Million cell updates/sec

Title: US-09-536-087-11
Perfect score: 32
Sequence: 1 CSVTVG 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	90.6	4302	3	US-08-658-136-5
2	29	90.6	4303	3	US-08-488-273-2
3	28	87.5	2161	3	US-09-081-320-3
4	28	87.5	3135	1	US-08-323-170B-2
5	27	84.4	6	1	US-07-801-812A-4
6	27	84.4	6	1	US-07-883-659-1
7	27	84.4	6	1	US-08-483-434A-1
8	27	84.4	6	1	US-08-487-568-4
9	27	84.4	6	2	US-08-858-971-4
10	27	84.4	6	3	US-08-971-538-1
11	27	84.4	6	4	US-08-476-134A-1
12	27	84.4	6	4	US-08-476-134A-31
13	27	84.4	6	6	5190920-3
14	27	84.4	6	6	5508208-4
15	27	84.4	6	6	5508208-37
16	27	84.4	7	1	US-08-483-434A-25
17	27	84.4	7	2	US-08-858-971-3
18	27	84.4	9	1	US-07-646-531D-1
19	27	84.4	9	1	US-07-646-531D-2
20	27	84.4	9	1	US-07-646-531D-3
21	27	84.4	9	1	US-07-646-531D-9
22	27	84.4	9	1	US-07-646-531D-11
23	27	84.4	9	1	US-07-646-531D-15
24	27	84.4	9	1	US-07-646-531D-16
25	27	84.4	9	1	US-07-646-531D-17
26	27	84.4	9	1	US-07-646-531D-18
27	27	84.4	9	2	US-08-488-273-1

28	27	84.4	9	2	US-08-488-273-2	Sequence 2, Appli
29	27	84.4	9	2	US-08-488-273-3	Sequence 3, Appli
30	27	84.4	9	2	US-08-488-273-9	Sequence 9, Appli
31	27	84.4	9	2	US-08-488-273-11	Sequence 11, Appli
32	27	84.4	9	2	US-08-488-273-13	Sequence 13, Appli
33	27	84.4	9	2	US-08-488-273-14	Sequence 14, Appli
34	27	84.4	9	2	US-08-858-971-1	Sequence 1, Appli
35	27	84.4	9	2	US-08-858-971-2	Sequence 2, Appli
36	27	84.4	9	2	US-08-318-856A-18	Sequence 18, Appli
37	27	84.4	9	2	US-08-318-856A-47	Sequence 47, Appli
38	27	84.4	9	6	5426100-1	Patent No. 5426100
39	27	84.4	9	6	5426100-2	Patent No. 5426100
40	27	84.4	9	6	5426100-3	Patent No. 5426100
41	27	84.4	9	6	5426100-9	Patent No. 5426100
42	27	84.4	11	1	US-08-483-434A-36	Sequence 36, Appli
43	27	84.4	11	4	US-08-476-134A-11	Sequence 11, Appli
44	27	84.4	12	2	US-08-858-971-7	Sequence 7, Appli
45	27	84.4	18	1	US-08-395-602A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-658-136-5
; Sequence 5, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,136
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: GEN4-17.8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-658-136-5

Query Match 90.6%; Score 29; DB 3; Length 4302;
Best Local Similarity 83.3%; Pred. No. 1.6e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 CSVTVG 6
    I:||||
Db 1270 CTVTVG 1275

RESULT 2
US-08-460-751-2
; Sequence 2, Application US/08460751
; Patent No. 5891628
; GENERAL INFORMATION:
; APPLICANT: Reeder, Stephen
; APPLICANT: Schneider, Michael
; APPLICANT: Glucksmann, Sandra
; TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
; TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,751
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/413,580
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7638-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4303 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-460-751-2

Query Match 90.68; Score 29; DB 2; Length 4303;
Best Local Similarity 83.36; Pred. No. 1.6e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CSVTVG 6
    I:||||
Db 1270 CTVTVG 1275

RESULT 3
US-09-081-320-3
; Sequence 3, Application US/09081320
; Patent No. 6093544
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Meng, Baozhong
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP

```

```

; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/081,320
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/047,147
; FILING DATE: 20-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/069,902
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1722
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2161 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-081-320-3

Query Match 87.5%; Score 28; DB 3; Length 2161;
Best Local Similarity 66.7%; Pred. No. 1.3e+03;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CSVTVG 6
    I:||||
Db 837 CSLTIG 842

RESULT 4
US-08-323-170B-2
; Sequence 2, Application US/08323170B
; Patent No. 5733772
; GENERAL INFORMATION:
; APPLICANT: Williamson, Kim C.
; APPLICANT: Kaslow, David C.
; TITLE OF INVENTION: Cloning and Expression of Plasmodium
; TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pfs230
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,170B
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:

```



```

; APPLICATION NUMBER: US 08/010,409
; FILING DATE: 29-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Quine, Jonathan A.
; REGISTRATION NUMBER: P-41,261
; REFERENCE/DOCKET NUMBER: 015280-1131000S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-323-170B-2

Query Match      87.5%; Score 28; DB 1; Length 3135;
Best Local Similarity 66.7%; Pred. No. 1.8e+03;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 CSVTVG 6
DB      3010 CNVTIG 3015

RESULT 5
US-07-801-812A-4
; Sequence 4, Application US/07801812A
; Patent No. 5357041
; GENERAL INFORMATION:
; APPLICANT: David D. Roberts et al
; TITLE OF INVENTION: HEPARIN-AND SULFATIDE-BINDING
; TITLE OF INVENTION: PEPTIDES FROM THE TYPE I REPEATS OF HUMAN
; TITLE OF INVENTION: THROMBOSPONDIN PROMOTE MELANOMA CELL ADHESION
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lowe, Price, LeBlanc & Becker
; STREET: Suite 300, 99 Canal Center Plaza
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: DOS TEXT FILE
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/801,812A
; FILING DATE: 19911206
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: J.G. Mullins
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 717-111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 684 1111
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-801-812A-4

Query Match      84.4%; Score 27; DB 1; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CSVTVG 6
DB      3010 CNVTIG 3015

APPLICATION NUMBER: US 08/010,409
FILING DATE: 29-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 015280-1131000S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3135 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-323-170B-2

Query Match      87.5%; Score 28; DB 1; Length 3135;
Best Local Similarity 66.7%; Pred. No. 1.8e+03;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 CSVTVG 6
DB      3010 CNVTIG 3015

RESULT 6
US-07-883-659-1
; Sequence 1, Application US/07883659
; Patent No. 5367059
; GENERAL INFORMATION:
; APPLICANT: Tuszyński, George Paul
; APPLICANT: Eyal, Jacob
; APPLICANT: Hamilton, Bruce King
; TITLE OF INVENTION: PURIFICATION AND CHARACTERIZATION
; TITLE OF INVENTION: OF A CSVTGC-SPECIFIC TUMOR CELL
; TITLE OF INVENTION: ADHESION RECEPTOR
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. R. Grace & Co.-Conn.
; STREET: 7379 Route 32
; CITY: Columbia
; STATE: Maryland
; COUNTRY: USA
; ZIP: 21044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/883,659
; FILING DATE: 19920514
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Appleby, Vanessa L.
; REGISTRATION NUMBER: 33223
; REFERENCE/DOCKET NUMBER: 01-8110
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 531-4515
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-883-659-1

Query Match      84.4%; Score 27; DB 1; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CSVTVG 6
DB      1 CSVTGC 6

RESULT 7
US-08-483-434A-1
; Sequence 113, Application US/08390156A
; Patent No. 5648458
; GENERAL INFORMATION:
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Dower, William J.
; APPLICANT: Martens, Christine L.
; TITLE OF INVENTION: Peptides and Compounds That Bind to
; TITLE OF INVENTION: ELAM-1
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Affymax Technologies, N.V.
; STREET: 4001 Miranda Ave.
; CITY: Palo Alto
; STATE: California
```

COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/390,156A
FILING DATE: 16-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/057,295
FILING DATE: 05-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/881,395
FILING DATE: 06-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Swiss, Gerald F.
REGISTRATION NUMBER: 30,113
REFERENCE/DOCKET NUMBER: 1023.1A
TELEPHONE: 415-496-2300
TELEFAX: 415-424-0832
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Region
LOCATION: 1
OTHER INFORMATION: /note= "Xaa is selected
OTHER INFORMATION: from the group consisting of His, Glu or Asp."
FEATURE:
NAME/KEY: Region
LOCATION: 2
OTHER INFORMATION: /note= "Xaa is selected
OTHER INFORMATION: from the group consisting of Ile, Met or Nle."
FEATURE:
NAME/KEY: Region
LOCATION: 3
OTHER INFORMATION: /note= "Xaa is selected
OTHER INFORMATION: from the group consisting of Thr or Ser."
FEATURE:
NAME/KEY: Region
LOCATION: 4
OTHER INFORMATION: /note= "Xaa is selected
OTHER INFORMATION: from the group consisting of Asp, Glu or Leu."
FEATURE:
NAME/KEY: Region
LOCATION: 5
OTHER INFORMATION: /note= "Xaa is selected
OTHER INFORMATION: from the group consisting of Gln or Glu."
FEATURE:
NAME/KEY: Region
LOCATION: 6
OTHER INFORMATION: /note= "Xaa is selected
OTHER INFORMATION: from the group consisting of Asn or Asp."
FEATURE:
NAME/KEY: Region
LOCATION: 7
OTHER INFORMATION: /note= "Xaa is selected
OTHER INFORMATION: from the group consisting of Leu, Met, Val or Ile."
FEATURE:
NAME/KEY: Region
LOCATION: 8
OTHER INFORMATION: /note= "Xaa is selected
OTHER INFORMATION: from the group consisting of Met or Nle."
FEATURE:
NAME/KEY: Region
LOCATION: 9

OTHER INFORMATION: /note= "Xaa is selected
OTHER INFORMATION: from the group consisting of Asn, Ser or Gln."
Sequence 1, Application: US/08483434A
Patent No. 5648461
GENERAL INFORMATION:
APPLICANT: EVAL, Jacob
APPLICANT: TUSZYNSKI, Bruce K.
APPLICANT: TUSZYNSKI, George P.
TITLE OF INVENTION: Synthetic Analogs of Thrombospondin and
TITLE OF INVENTION: Therapeutic Use Thereof
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: 1601 Market Street, 36th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103-2398
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,434A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/450,738
FILING DATE: 25-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/185,614
FILING DATE: 24-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/024,436
FILING DATE: 01-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/587,197
FILING DATE: 24-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/483,527
FILING DATE: 22-FEB-1990
ATTORNEY/AGENT INFORMATION:
NAME: Leary Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9598-306 (9049)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 567-2020
TELEFAX: (215) 567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-483-434A-1

Query Match 84.4%; Score 27; DB 1; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CSVTVG 6
Db 1 CSVTCG 6

RESULT 8
US-08-487-568-4
; Sequence 4, Application US/08487568
; Patent No. 5770563

```
;
; GENERAL INFORMATION:
; APPLICANT: Roberts, David D.
; APPLICANT: Browning, Philip J.
; APPLICANT: Bryant, Joseph L.
; APPLICANT: Inman, John K.
; APPLICANT: Krutzsch, Henry C.
; APPLICANT: Guo, Nenghua
; TITLE OF INVENTION: Heparin and Sulfatide Binding Peptides
; TITLE OF INVENTION: from the Type-I Repeats of Human Trombospondin and
; TITLE OF INVENTION: Conjugates Thereof
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,568
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/215,085
; FILING DATE: 21-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/801,812
; FILING DATE: 06-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 015280-023310
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-487-568-4

Query Match 84.4%; Score 27; DB 1; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 1 CSVTGC 6

RESULT 9
US-08-858-971-4
; Sequence 4, Application US/08858971
; Patent No. 5886142
; GENERAL INFORMATION:
; APPLICANT: M. Thakur, V. Pallela
; TITLE OF INVENTION: Radiolabeled Thrombus Imaging Agents
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jane Massey Licata, Esq.
; STREET: 210 Lake Drive East, Suite 201
; CITY: Cherry Hill
; STATE: NJ
; COUNTRY: USA

Query Match 84.4%; Score 27; DB 3; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 1 CSVTGC 6

US-08-971-538-1
; Sequence 1, Application US/08971538
; Patent No. 6090367
; GENERAL INFORMATION:
; APPLICANT: Khalil, Nasreen
; TITLE OF INVENTION: Post-Translational Activation of TGF-beta 1 Involving
; FILE REFERENCE: 916.6607P
; CURRENT APPLICATION NUMBER: US/08/971,538
; CURRENT FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: GB 9509957.8
; EARLIER FILING DATE: 1995-05-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope of
; OTHER INFORMATION: thrombospondin-1
; US-08-971-538-1

Query Match 84.4%; Score 27; DB 3; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 1 CSVTGC 6

RESULT 11
US-08-476-134A-1
```

```
;
; ZIP: 08002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,971
; FILING DATE: herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane Massey Licata
; REGISTRATION NUMBER: 32,257
; REFERENCE/DOCKET NUMBER: JEFF-0196
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 779-2400
; TELEFAX: (609) 779-8488
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-858-971-4

Query Match 84.4%; Score 27; DB 2; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 1 CSVTGC 6

RESULT 10
US-08-971-538-1
; Sequence 1, Application US/08971538
; Patent No. 6090367
; GENERAL INFORMATION:
; APPLICANT: Khalil, Nasreen
; TITLE OF INVENTION: Post-Translational Activation of TGF-beta 1 Involving
; FILE REFERENCE: 916.6607P
; CURRENT APPLICATION NUMBER: US/08/971,538
; CURRENT FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: GB 9509957.8
; EARLIER FILING DATE: 1995-05-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope of
; OTHER INFORMATION: thrombospondin-1
; US-08-971-538-1

Query Match 84.4%; Score 27; DB 3; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 1 CSVTGC 6

RESULT 11
US-08-476-134A-1
```

```
; Sequence 1, Application US/08476134A
; Patent No. 6239110
; GENERAL INFORMATION:
; APPLICANT: EYAL, JACOB
; APPLICANT: HAMILTON, BRUCE K.
; APPLICANT: TUSZYNSKI, GEORGE P.
; TITLE OF INVENTION: SYNTHETIC ANALOGS OF THROMBOSPONDIN AND THERAPEUTIC USE
; TITLE OF INVENTION: THEROF
; FILE REFERENCE: 07206-0009
; CURRENT APPLICATION NUMBER: US/08/476,134A
; CURRENT FILING DATE: 1995-06-07
; PRIOR FILING DATE: 07/587,197
; PRIOR APPLICATION NUMBER: 07/483,527
; PRIOR FILING DATE: 1990-09-24
; PRIOR APPLICATION NUMBER: 08/450,738
; PRIOR FILING DATE: 1995-05-25
; PRIOR APPLICATION NUMBER: 08/185,614
; PRIOR FILING DATE: 1994-01-24
; PRIOR APPLICATION NUMBER: 08/024,436
; PRIOR FILING DATE: 1993-03-01
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: analog of thrombospondin
; OTHER INFORMATION: optionally disulfide linked
; US-08-476-134A-1

Query Match      84.4%; Score 27; DB 4; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 1 CSVTGC 6

RESULT 12
US-08-476-134A-31
; Sequence 31, Application US/08476134A
; Patent No. 6239110
; GENERAL INFORMATION:
; APPLICANT: EYAL, JACOB
; APPLICANT: HAMILTON, BRUCE K.
; APPLICANT: TUSZYNSKI, GEORGE P.
; TITLE OF INVENTION: SYNTHETIC ANALOGS OF THROMBOSPONDIN AND THERAPEUTIC USE
; TITLE OF INVENTION: THEROF
; FILE REFERENCE: 07206-0009
; CURRENT APPLICATION NUMBER: US/08/476,134A
; CURRENT FILING DATE: 1995-06-07
; PRIOR FILING DATE: 07/587,197
; PRIOR APPLICATION NUMBER: 07/483,527
; PRIOR FILING DATE: 1990-09-24
; PRIOR APPLICATION NUMBER: 08/450,738
; PRIOR FILING DATE: 1995-05-25
; PRIOR APPLICATION NUMBER: 08/185,614
; PRIOR FILING DATE: 1994-01-24
; PRIOR APPLICATION NUMBER: 08/024,436
; PRIOR FILING DATE: 1993-03-01
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
```

```
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: analog of thrombospondin
; US-08-476-134A-31

Query Match      84.4%; Score 27; DB 4; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 1 CSVTGC 6

RESULT 13
5190920-3
; Patent No. 5190920
; APPLICANT: EYAL, JACOB; HAMILTON, BRUCE K.; TUSZYNSKI,
; GEORGE P.
; TITLE OF INVENTION: METHOD FOR USING SYNTHETIC ANALOGS
; OF THROMBOSPONDIN FOR INHIBITING METASTASIS ACTIVITY
; NUMBER OF SEQUENCES: 32
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/587,197
; FILING DATE: 24-SEP-1990
; SEQ ID NO: 3
; LENGTH: 6
; 5190920-3

Query Match      84.4%; Score 27; DB 6; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 1 CSVTGC 6

RESULT 14
5506208-4
; Patent No. 5506208
; APPLICANT: EYAL, JACOB; HAMILTON, BRUCE K.; TUSZYNSKI,
; GEORGE P.
; TITLE OF INVENTION: METHOD FOR USING SYNTHETIC ANALOGS OF
; THROMBOSPONDIN FOR INHIBITING ANGIOGENESIS ACTIVITY
; NUMBER OF SEQUENCES: 45
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/408,181
; FILING DATE: 22-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 131,565
; FILING DATE: 04-OCT-1993
; APPLICATION NUMBER: 895,764
; FILING DATE: 09-JUN-1992
; APPLICATION NUMBER: 587,197
; FILING DATE: 24-SEP-1990
; SEQ ID NO: 4
; LENGTH: 6
; 5506208-4

Query Match      84.4%; Score 27; DB 6; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 1 CSVTGC 6

RESULT 15
5506208-37
```

; Patent No. 5506208
; APPLICANT: EYAL, JACOB; HAMILTON, BRUCE K.; TUSZYNSKI,
; GEORGE P.
; TITLE OF INVENTION: METHOD FOR USING SYNTHETIC ANALOGS OF
; THROMBOSPONDIN FOR INHIBITING ANGIOGENESIS ACTIVITY
; NUMBER OF SEQUENCES: 45
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/408,181
; FILING DATE: 22-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 131,565
; FILING DATE: 04-OCT-1993
; APPLICATION NUMBER: 895,764
; FILING DATE: 09-JUN-1992
; APPLICATION NUMBER: 587,197
; FILING DATE: 24-SEP-1990
; SEQ ID NO:37:
; LENGTH: 6
; 5506208-37

Query Match 84.4%; Score 27; DB 6; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVTVG 6
 |||||
Db 1 CSVTCG 6

Search completed: January 31, 2002, 09:23:58
Job time: 223 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:36:19 ; Search time 90.11 seconds
(without alignments)
9.740 Million cell updates/sec

Title: US-09-536-087-11

Perfect score: 32

Sequence: 1 CSVTG 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	100.0	596	11 Q99LB8	Q99LB8 mus musculus
2	32	100.0	774	5 Q9GP19	Q9GP19 helioidarid
3	32	100.0	780	5 Q9GP17	Q9GP17 helioidarid
4	32	100.0	1192	10 Q957T0	Q957T0 arabidopsis
5	31	96.9	196	10 Q9LJL5	Q9LJL5 arabidopsis
6	31	96.9	495	3 Q96858	Q96858 saccharomyc
7	29	90.6	104	2 Q32957	Q32957 mycobacteri
8	29	90.6	134	12 Q98245	Q98245 molluscum c
9	29	90.6	138	5 Q9T251	Q9T251 caenorhabdi
10	29	90.6	357	5 Q9XV87	Q9XV87 caenorhabdi
11	29	90.6	392	5 Q9XAP6	Q9XAP6 caenorhabdi
12	29	90.6	440	2 Q9KVS7	Q9KVS7 vibrio chol
13	29	90.6	479	12 Q84808	Q84808 puma lentiv
14	29	90.6	538	2 Q9I3R9	Q9I3R9 pseudomonas
15	29	90.6	870	10 Q9FJD4	Q9FJD4 arabidopsis
16	29	90.6	942	3 Q9UV86	Q9UV86 botryosphae
17	29	90.6	1345	2 Q9LO60	Q9LO60 streptomyce
18	29	90.6	1429	5 Q9V3C5	Q9V3C5 drosophila
19	29	90.6	3638	4 Q15142	Q15142 homo sapien

20	29	90.6	4292	4 Q15141	Q15141 homo sapien
21	29	90.6	4302	4 Q15140	Q15140 homo sapien
22	28	87.5	48	10 P93388	P93388 nicotiana t
23	28	87.5	101	4 Q9H3S9	Q9H3S9 homo sapien
24	28	87.5	101	4 Q9G2S4	Q9G2S4 homo sapien
25	28	87.5	144	10 Q41912	Q41912 arabidopsis
26	28	87.5	158	2 P73236	P73236 synecocyst
27	28	87.5	185	8 Q9GFI9	Q9GFI9 illicium pa
28	28	87.5	188	8 Q9GFJ3	Q9GFJ3 cercidiphyll
29	28	87.5	197	8 Q9GFJ6	Q9GFJ6 saurus ce
30	28	87.5	199	8 Q9GFJ5	Q9GFJ5 calycanthus
31	28	87.5	200	8 Q9GFJ6	Q9GFJ6 cabomba car
32	28	87.5	201	8 Q9MSP3	Q9MSP3 nympheae od
33	28	87.5	201	8 Q9GFJ8	Q9GFJ8 acorus cala
34	28	87.5	201	8 Q9GFJ7	Q9GFJ7 asarum cana
35	28	87.5	201	8 Q9GFJ4	Q9GFJ4 ceratophyll
36	28	87.5	201	8 Q9GFJ8	Q9GFJ8 lactoris fe
37	28	87.5	201	8 Q9GFJ7	Q9GFJ7 liriodendro
38	28	87.5	201	8 Q9GFJ5	Q9GFJ5 trochodendr
39	28	87.5	201	8 Q9GE23	Q9GE23 amborella t
40	28	87.5	220	2 Q9JUA6	Q9JUA6 neisseria m
41	28	87.5	234	2 Q9J282	Q9J282 neisseria m
42	28	87.5	250	2 Q07720	Q07720 mycobacteri
43	28	87.5	272	8 Q9THV0	Q9THV0 spinacia ol
44	28	87.5	272	10 Q9M3I1	Q9M3I1 spinacia ol
45	28	87.5	274	8 Q9MDU0	Q9MDU0 oenothera h

ALIGNMENTS

RESULT 1
Q99LB8
ID Q99LB8 PRELIMINARY; PRT; 596 AA.
AC Q99LB8;
DT 01-JUN-2001 (TREMREL. 17, Created)
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE SIMILAR TO FKBP-ASSOCIATED PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003446; AAH03446.1; -;
SQ SEQUENCE 596 AA; 67679 MW; 9E986E74F6A7CFFA CRC64;

Query Match 100.0%; Score 32; DB 11; Length 596;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTG 6
| | | | |
DB 539 CSVTG 544

RESULT 2
Q9GP19
ID Q9GP19 PRELIMINARY; PRT; 774 AA.
AC Q9GP19;
DT 01-MAR-2001 (TREMREL. 16, Created)
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
DE MSP130 PROTEIN.
GN MSP130.
OS Heliocidaris erythrogramma (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Echinometridae;
OC Heliocidaris.

```
OX NCBI_TaxID=7634;
RN [1]
RP SEQUENCE FROM N.A.
RA Klueg K.M., Raff R.A.;
RT "Structure and evolution of a heterochronically expressed gene in the
RL direct-developing sea urchin, Helicodiaris erythrogramma.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ278115; CAC20358.1; -.
SQ SEQUENCE 774 AA; 81819 MW; 8C34E99E720149BA CRC64;

Query Match 100.0%; Score 32; DB 5; Length 774;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 151 CSVTVG 156
|||||

RESULT 3
Q9GP17 PRELIMINARY; PRT; 780 AA.
AC Q9GP17;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE MSPI30 PROTEIN.
GN MSPI30.
OS Helicodiaris tuberculata (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Echinometridae;
OC Helicodiaris.
OX NCBI_TaxID=7635;
RN [1]
RP SEQUENCE FROM N.A.
RA Klueg K.M., Raff R.A.;
RT "Structure and evolution of a heterochronically expressed gene in the
RL direct-developing sea urchin, Helicodiaris erythrogramma.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ278114; CAC20589.1; -.
SQ SEQUENCE 780 AA; 82135 MW; FBD99DB412D81559 CRC64;

Query Match 100.0%; Score 32; DB 5; Length 780;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 157 CSVTVG 162
|||||

RESULT 4
Q9S7T0 PRELIMINARY; PRT; 1192 AA.
AC Q9S7T0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE F28J7.11 PROTEIN.
GN F28J7.11 OR F4P13.33.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
SQ SEQUENCE 774 AA; 81819 MW; 8C34E99E720149BA CRC64;
```

```
RT "Arabidopsis thaliana chromosome III BAC F28J7 genomic sequence.";
RN Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC F4P13 genomic sequence.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC010797; AAF03433.1; -.
DR EMBL; AC009325; AAF01560.1; -.
DR InterPro; IPR002048; EF-hand.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
SQ SEQUENCE 1192 AA; 132863 MW; 4F67B124CBAFF154 CRC64;

Query Match 100.0%; Score 32; DB 10; Length 1192;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 861 CSVTVG 866
|||||

RESULT 5
Q9LJL5 PRELIMINARY; PRT; 196 AA.
ID Q9LJL5
AC Q9LJL5;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE GBIAC23758.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
DR EMBL; AP000419; BAB02955.1; -.
SQ SEQUENCE 196 AA; 21454 MW; AC543304CE6B92C4 CRC64;

Query Match 96.9%; Score 31; DB 10; Length 196;
Best Local Similarity 83.3%; Pred. No. 15;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 59 CSVTVG 64
|||||

RESULT 6
Q06858 PRELIMINARY; PRT; 495 AA.
ID Q06858
AC Q06858;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
```


DE PHO2 PROTEIN.
GN PHO2 OR YDL106C.
OS Saccharomyces cerevisiae (Baker's yeast).
NC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GG100-14D;
RA Bergman L.W.;
RL Submitted (AUG-1990) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-184 FROM N.A.
RX STRAIN=GG100-14D;
RC MEDLINE=91288241; PubMed=2062666;
RA McCarthy B.J., Creasy C.L., Bergman L.W.;
RL "Molecular analysis of a temperature sensitive allele of the PHO2 gene
of Saccharomyces cerevisiae.";
RT Nucleic Acids Res. 19:3463-3463(1991).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER HOMEBOX DOMAINS.
DR EMBL: X54293; CAA38192.1; --
DR SGD: S0002264; PHO2
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
DR DNA-binding; Homeobox; Nuclear protein.
KW
SQ SEQUENCE 495 AA; 55527 MW; 781D3A5B0E6E6F2D CRC64;

Query Match 96.9%; Score 31; DB 3; Length 495;
Best Local Similarity 83.3%; Pred. No. 36;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CSVTVG 6
I:||||
Db 136 CSITVG 141

RESULT 7
ID O32957 PRELIMINARY; PRT; 104 AA.
AC O32957;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 11.1 KDA PROTEIN (HYPOTHETICAL PROTEIN ML0863).
GN MLCB22.15C OR ML0863.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Devlin K., Churcher C.M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=93188700; PubMed=8446027;
RA Eglmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization
of Mycobacterium leprae.";
RL Mol. Microbiol. 7:197-206(1993).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;

RA Cole S.T., Eglmeier K., Parkhill J., James K.D., Thomson N.R.,
Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RL "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL: Z98741; CAB11380.1; --
DR EMBL: AL583920; CAC31244.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 104 AA; 11055 MW; C8E2D845984B5451 CRC64;

Query Match 90.6%; Score 29; DB 2; Length 104;
Best Local Similarity 83.3%; Pred. No. 24;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CSVTVG 6
I:||||
Db 82 CTVTVG 87

RESULT 8
ID Q98245 PRELIMINARY; PRT; 134 AA.
AC Q98245;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE MC078L.
GN MC078L.
OS Molluscum contagiosum virus subtype 1 (MCV1).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Molluscipoxvirus.
OX NCBI_TaxID=10280;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96325459; PubMed=8670425;
RA Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,
RA Moss B.;
RT "Genome sequence of a human tumorigenic poxvirus: prediction of
specific host response-evasion genes.";
RL Science 273:813-816(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,
RA Moss B.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U60315; AAC55206.1; --
SQ SEQUENCE 134 AA; 14732 MW; 1BEC5DD53EA9182D CRC64;

Query Match 90.6%; Score 29; DB 12; Length 134;
Best Local Similarity 83.3%; Pred. No. 31;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CSVTVG 6
I:||||
Db 72 CSVTIG 77

RESULT 9
ID Q9TZ51 PRELIMINARY; PRT; 138 AA.
AC Q9TZ51;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE K01A2.2 PROTEIN.


```
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE THIAMIN-PHOSPHATE PYROPHOSPHORYLASE.
GN VC0062.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickley E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004097; AAF93240.1; -.
DR TIGR; VC0062; -.
DR InterPro; IPR003733; TMP_synthase.
DR Pfam; PF02581; TMP-TEN1; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 440 AA; 48902 MW; 0703E84D24BE83C4 CRC64;

Query Match 90.6%; Score 29; DB 2; Length 440;
Best Local Similarity 83.3%; Pred. No. 94;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
   |:||||
Db 92 CTVTG 97

RESULT 13
Q84808
ID Q84808 PRELIMINARY; PRT; 479 AA.
AC Q84808;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE GAG POLYPEPTIDE [CONTAINS: CORE PROTEINPRECURSOR(S) P24].
GN GAG.
OS Puma lentivirus 14.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=32615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94303200; PubMed=8030248;
RA Langley R.J., Hirsch V.M., O'Brien S.J., Adger-Johnson D.,
RA Goeken R.M., Olmsted R.A.;
RT "Nucleotide sequence analysis of puma lentivirus (PLV-14): genomic
RT organization and relationship to other lentiviruses.";
RL Virology 202:853-864(1994).
CC -!- FUNCTION: THE CCHC ZINC FINGER DOMAIN IS REQUIRED FOR VIRAL GENOME
CC -!- PACKAGING AND FOR EARLY INFECTION PROCESS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO ZN-FINGER CCHC TYPE FAMILY.
CC -!- SIMILARITY: TO GAG GENE PROTEIN P24 (CORE NUCLEOCAPSID PROTEIN).
DR EMBL; U03982; AAA67167.1; -.
DR HSSP; P03351; 1E1A.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00607; gag_p24; 1.
DR PRINTS; PR00098; zf-CCHC; 2.
DR Pfam; PR00939; C2HCZNFINGER.
DR SMART; SM00343; Znf_C2HC; 2.
KW Core protein; Polyprotein; Zinc-finger.
SQ SEQUENCE 479 AA; 52647 MW; 21C4190013B21242 CRC64;
```

```
Query Match 90.6%; Score 29; DB 2; Length 538;
Best Local Similarity 83.3%; Pred. No. 1.le+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
   |:||||
Db 397 CAVTVG 402

RESULT 15
Q9FJD4
ID Q9FJD4 PRELIMINARY; PRT; 870 AA.
AC Q9FJD4;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE IMPORTIN BETA.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=99087489; PubMed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
```

```
Query Match 90.6%; Score 29; DB 12; Length 479;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
   |:||||
Db 17 CNVTVG 22

RESULT 14
Q913R9
ID Q913R9 PRELIMINARY; PRT; 538 AA.
AC Q913R9;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PROBABLE ATP-BINDING COMPONENT OF ABC TRANSPORTER.
GN PA1425.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004572; AAG04814.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 538 AA; 59072 MW; 5E9A4835CB2C7EE3 CRC64;
```

RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT Sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned pl and TAC clones.";
RL DNA Res. 5:297-308(1998).
DR EMBL; AB015476; BAB09724.1; -.
DR InterPro; IPR001494; IBN_NT.
SQ SEQUENCE 870 AA; 96258 MW; E09C2DF6980889B CRC64;

Query Match 90.6%; Score 29; DB 10; Length 870;
Best Local Similarity 83.3%; Pred. NO. 1.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CSVTVG 6
I:||||
Db 665 CAVTVG 670

Search completed: January 31, 2002, 09:36:20
Job time: 789 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:36:52 ; Search time 28.31 Seconds
(without alignments)
7.771 Million cell updates/sec

Title: US-09-536-087-11
Perfect score: 32
Sequence: 1 CSVTVG 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	31	96.9	559	1 PHO2_YEAST	P07269 saccharomyc
2	29	90.6	207	1 SNAT_MESAU	Q9r089 mesocricetu
3	29	90.6	410	1 Y474_AQUAE	O66772 aquifex ao
4	29	90.6	535	1 ICAL_BOVIN	O95132 bos taurus
5	29	90.6	695	1 YD38_YEAST	Q05497 saccharomyc
6	29	90.6	705	1 Y4S0_RHLSN	P55656 rhizobium s
7	29	90.6	4303	1 PKD1_HUMAN	P98161 homo sapien
8	28	87.5	143	1 RK2_SOYBN	P18663 glycine max
9	28	87.5	204	1 SNAT_MACMU	O97756 macaca mula
10	28	87.5	205	1 SNAT_CHICK	P79774 gallus gall
11	28	87.5	205	1 SNAT_MOUSE	O88816 mus musculu
12	28	87.5	207	1 SNAT_HUMAN	Q16613 homo sapien
13	28	87.5	229	1 YXHK_CAEEL	Q20728 caenorhabdi
14	28	87.5	266	1 RK2_NICDE	P21434 nicotiana d
15	28	87.5	269	1 NIVA_CLOPA	Q00853 clostridium
16	28	87.5	271	1 RK2_SPTOL	P06509 spinacia ol
17	28	87.5	273	1 RK2_PEA	P31163 pisum sativ
18	28	87.5	274	1 RK2_ARATH	P56791 arabidopsis
19	28	87.5	274	1 RK2_EPIVI	P30065 epifagus vi
20	28	87.5	274	1 RK2_OENHO	Q9mdu0 oenothera h
21	28	87.5	274	1 RK2_SINAL	P27107 sinapis alb
22	28	87.5	274	1 RK2_TOBAC	P06379 nicotiana t
23	28	87.5	333	1 OM1K_CHLPN	Q9xbf4 chlamydia p
24	28	87.5	389	1 OM1N_CHLPN	Q07430 chlamydia p
25	28	87.5	389	1 OMPL_CHLPN	P27455 chlamydia p
26	28	87.5	3135	1 S230_PLAFO	Q08372 plasmodium
27	27	84.4	272	1 RL2_ACTAC	P55835 actinobacil
28	27	84.4	275	1 RK2_PICAB	O62954 picea abies
29	27	84.4	276	1 RK2_PINTH	O62940 pinus thunb
30	27	84.4	343	1 CSP_PLAVS	P13826 plasmodium
31	27	84.4	351	1 CSP_PLAKH	P04922 plasmodium
32	27	84.4	363	1 CSP_PLAHO	P02894 plasmodium
33	27	84.4	367	1 CSP_PLAYO	P06914 plasmodium

34	27	84.4	378	1 CSP_PLACB	P08672 plasmodium
35	27	84.4	378	1 CSP_PLAVB	P08677 plasmodium
36	27	84.4	386	1 CSP_PLAST	Q03110 plasmodium
37	27	84.4	388	1 CSP_PLARE	P26694 plasmodium
38	27	84.4	393	1 CSP_PLABR	P14593 plasmodium
39	27	84.4	397	1 CSP_PLAFO	P19597 plasmodium
40	27	84.4	398	1 CSP_PLACC	P08673 plasmodium
41	27	84.4	412	1 CSP_PLAFA	P02893 plasmodium
42	27	84.4	419	1 CSP_PLACM	P08676 plasmodium
43	27	84.4	424	1 CSP_PLAFT	P13814 plasmodium
44	27	84.4	429	1 CSP_PLAWA	P13815 plasmodium
45	27	84.4	437	1 PROP_MOUSE	P11680 mus musculu

ALIGNMENTS

RESULT 1
PHO2_YEAST
ID PHO2_YEAST STANDARD; PRT; 559 AA.
AC P07269;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE REGULATORY PROTEIN PHO2 (GENERAL REGULATORY FACTOR 10).
GN PHO2 OR BAS2 OR GRF10 OR YDL106C OR D2350.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87146334; PubMed=3029672;
RA Sengstag C., Hinnen A.;
RT "The sequence of the Saccharomyces cerevisiae gene PHO2 codes for a
regulatory protein with unusual aminoacid composition.";
RL Nucleic Acids Res. 15:233-246(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA Berben G.;
RL Submitted (JAN-1989) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97051597; PubMed=8896274;
RA Saiz J.E., Buitrago M.J., Garcia R., Revuelta J.L., del Rey F.;
RT "The sequence of a 20.3 kb DNA fragment from the left arm of
Saccharomyces cerevisiae chromosome IV contains the KIN28, MSS2,
PHO2, POL3 and DUN1 genes, and six new open reading frames.";
RL Yeast 12:1077-1084(1996).
RN [4]
RP STRUCTURE, EXPRESSION, AND FUNCTION.
RX MEDLINE=89006272; PubMed=3049251;
RA Berben G., Legrain M., Hilger F.;
RT "Studies on the structure, expression and function of the yeast
regulatory gene PHO2.";
RL Gene 66:307-312(1988).
RN [5]
RP HOMEBOX DOMAIN.
RX MEDLINE=88210458; PubMed=2896548;
RA Buerklin T.R.;
RT "The yeast regulatory gene PHO2 encodes a homeo box.";
RL Cell 53:339-340(1988).
CC -1- FUNCTION: REGULATOR IN PHOSPHATE METABOLISM AND ACTS AS A
DEREPRESSOR OF ANOTHER CENTRAL REGULATOR PHOS. BINDS TO THE
UPSTREAM ACTIVATOR SEQUENCE (UAS) OF PHO5. IT ALSO BINDS TO THE
TRP4, HIS4, AND CYC1 PROMOTERS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: TWO REGIONS OF STRONG HOMOLOGY TO PHO4 ARE FOUND.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; X05062; CAA28729.1; -
DR EMBL; M22259; AAA34866.1; -
DR EMBL; X95644; CAA64906.1; -
DR EMBL; 274154; CAA88673.1; -
DR PIR; A25872; A25872.
DR HSSP; P02836; 1HDD.
DR TRANSFAC; T00689; -
DR SGD; S0002264; GRF10.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Transcription regulation; Nuclear protein.
FT DOMAIN 23 52 GLN-RICH (INVOLVED IN TRANSCRIPTIONAL
FT ACTIVATION) (POTENTIAL).
FT DNA_BIND 77 136 HOMEBOX.
SQ SEQUENCE 559 AA; 63390 MW; BE20E396D6AA0281 CRC64;

Query Match 96.9%; Score 31; DB 1; Length 559;
Best Local Similarity 83.3%; Pred. No. 12;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
||:||||
DB 200 CSITVG 205

RESULT 2

SNAT_MESAU
ID SNAT_MESAU STANDARD; PRT; 207 AA.
AC Q9ROA9;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SEROTONIN N-ACETYLTRANSFERASE (EC 2.3.1.87) (ARYALKYLAMINE N-
DE ACETYLTRANSFERASE) (AA-NAT) (SEROTONIN ACETYLASE).
OS AANAT OR SNAT.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pineal gland.
RX MEDLINE=99337801; PubMed=10407190;
RA Gauer F., Poirel V.J., Garidou M.L., Simonneaux V., Pevet P.;
RT Molecular cloning of the arylalkylamine-N-acetyltransferase and daily
RT variations of its mRNA expression in the syrian hamster pineal
RT gland*;
RL Brain Res. Mol. Brain Res. 71:87-95(1999).
CC -!- FUNCTION: CATALYZES THE N-ACETYLATION OF SEROTONIN INTO N-
CC ACETYLSEROTONIN.
CC -!- CATALYTIC ACTIVITY: ACETYL-COA + ARALKYLAMINE = COA + N-
CC ACETYLARALKYLAMINE.
CC -!- PATHWAY: CONVERSION OF SEROTONIN TO MELATONIN; PENULTIMATE STEP.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PINEAL GLAND AND RETINA.
CC -!- SIMILARITY: BELONGS TO THE ACETYLTRANSFERASE FAMILY. AANAT
CC SUBFAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AF092100; AAD55970.1; -
DR InterPro; IPR000182; Acetyltransf_GCN5.
DR Pfam; PF00583; Acetyltransf; 1.
DR Transferase; Acetyltransferase.
FT ACT_SITE 168 168 BY SIMILARITY.
SQ SEQUENCE 207 AA; 23385 MW; C179FC32F08C0734 CRC64;

Query Match 90.6%; Score 29; DB 1; Length 207;
Best Local Similarity 83.3%; Pred. No. 13;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
|:||||
DB 179 CAVTVG 184

RESULT 3

Y474_AQUAE
ID Y474_AQUAE STANDARD; PRT; 410 AA.
AC O66772;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN AQ_474.
GN AQ_474.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujoy M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus".;
RL Nature 392:353-358(1998).
CC -!- SIMILARITY: BELONGS TO THE UPF0004 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AE000691; AAC06733.1; -
DR InterPro; IPR001861; UPF00004.
DR Pfam; PF00919; UPF0004; 1.
DR PROSITE; PS01278; UPF0004; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 410 AA; 47833 MW; 62C6BCBF5E689E9B CRC64;

Query Match 90.6%; Score 29; DB 1; Length 410;
Best Local Similarity 83.3%; Pred. No. 25;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
|:||||
DB 46 CTVTVG 51

RESULT 4

ICAL_BOVIN
ID ICAL_BOVIN STANDARD; PRT; 535 AA.

Q95132;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
20-AUG-2001 (Rel. 40, Last annotation update)
INTERCELLULAR ADHESION MOLECULE-1 PRECURSOR (ICAM-1).
ICAM1.
Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=98100484; PubMed=9437830;
RA Lee E.K., Kang S.G., Kehrl M.E. Jr.;
RT "Cloning, sequencing and analysis of cDNA encoding bovine
intercellular adhesion molecule-1 (ICAM-1).";
RL Vet. Immunol. Immunopathol. 59:121-129(1997).
CC -!- FUNCTION: ICAM PROTEINS ARE LIGANDS FOR THE LEUKOCYTE ADHESION
MOLECULE-1 (INTEGRIN ALPHA-L/BETA-2).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE ICAM FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL: U65789; AAB06749.1; -
DR HSP: P05362; IIC1.
DR InterPro: IPR003599; Ig.
DR Pfam: PF00047; Ig_2.
DR SMART: SM00409; Ig_1.
KW Immunoglobulin domain; Cell adhesion; Glycoprotein; Transmembrane;
KW Repeat; Signal.
FT SIGNAL 1 27 BY SIMILARITY.
FT CHAIN 28 535 INTERCELLULAR ADHESION MOLECULE-1.
FT DOMAIN 28 479 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 481 503 POTENTIAL.
FT DOMAIN 504 535 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 41 102 IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 127 193 IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN 230 295 IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN 323 376 IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN 410 463 IG-LIKE C2-TYPE DOMAIN 5.
FT DISULFID 48 91 BY SIMILARITY.
FT DISULFID 52 95 BY SIMILARITY.
FT DISULFID 134 186 BY SIMILARITY.
FT DISULFID 237 288 BY SIMILARITY.
FT DISULFID 330 369 BY SIMILARITY.
FT DISULFID 417 456 BY SIMILARITY.
FT SITE 151 153 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 131 131 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 183 183 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 236 236 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 302 302 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 341 341 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 357 357 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 428 428 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 535 AA; 58428 MW; 7485EBE9C04D10B3 CRC64;

Query Match 90.6%; Score 29; DB 1; Length 535;
Best Local Similarity 83.3%; Pred. No. 33;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CSVTVG 6
DB 288 CSVTIG 293
|||||
RESULT 5
YD38_YEAST STANDARD; PRT; 695 AA.
AC Q05497;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOPHETICAL 77.8 KDA PROTEIN IN MRPS28-HXT7 INTERGENIC REGION.
GN YDR338C OR D9651.8.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favallo A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Marks E., Meneses S., Miller N., Nhan N., Pauley A., Peluso D.,
RA Rifken L., Riles L., Taich A., Trevasakis E., Vignati D.,
RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE MULTI ANTIMICROBIAL EXTRUSION
FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL: U51032; AAB64774.1; -
DR SGD: S0002746; YDR338C.
DR InterPro: IPR002528; UPF0013.
DR Pfam: PF01554; UPF0013; 2.
KW Hypothetical protein; Transmembrane; Transport.
FT TRANSMEM 237 257 POTENTIAL.
FT TRANSMEM 265 285 POTENTIAL.
FT TRANSMEM 313 333 POTENTIAL.
FT TRANSMEM 344 364 POTENTIAL.
FT TRANSMEM 380 400 POTENTIAL.
FT TRANSMEM 408 428 POTENTIAL.
FT TRANSMEM 457 477 POTENTIAL.
FT TRANSMEM 488 508 POTENTIAL.
FT TRANSMEM 531 551 POTENTIAL.
FT TRANSMEM 565 585 POTENTIAL.
FT TRANSMEM 604 624 POTENTIAL.
FT TRANSMEM 633 653 POTENTIAL.
SQ SEQUENCE 695 AA; 77846 MW; C6D95204FBEBFD6 CRC64;
Query Match 90.6%; Score 29; DB 1; Length 695;
Best Local Similarity 83.3%; Pred. No. 42;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CSVTVG 6
DB 253 CSLTVG 258
|||||

```

RESULT 6
Y4SO_RHISN
ID Y4SO_RHISN STANDARD; PRT; 705 AA.
AC P55656;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROBABLE PEPTIDASE Y4SO (EC 3.4.21.-).
GN Y4SO.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym pNGR234.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9A; ALSO KNOWN AS THE
CC PROLYL OLIGOPEPTIDASE FAMILY. STRONG, TO Y4QF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000097; AAB91854.1;
DR InterPro: IPR000379; Est_lip_thioest_actsite.
DR InterPro: IPR0001375; Peptidase_S9.
DR InterPro: IPR0002470; Proligo_Phase.
DR Pfam: PF00326; Peptidase_S9; 1.
DR PRINTS: PR00862; PROLIGOPHASE.
KW Hypothetical protein; Hydrolase; Serine protease; Plasmid.
FT ACT_SITE 554 554 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 676 676 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 705 AA; 79767 MW; EA4985F696A9880 CRC64;

Query Match 90.6%; Score 29; DB 1; Length 705;
Best Local Similarity 83.3%; Pred. NO. 43;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 369 CTVTVG 374
I:|||||

RESULT 7
PKD1_HUMAN
ID PKD1_HUMAN STANDARD; PRT; 4303 AA.
AC P98161;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE POLYCYSTIN PRECURSOR (AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY DISEASE
DE PROTEIN 1).
GN PKD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95254638; PubMed=7736581;
RX Gluecksmann-Kulis M.A., Tayber O., Woolf E.A., Bougueleret L.,
RA Deng N., Alperin G.D., Iris F., Hawkins F., Munro C., Lakey N.,

```

```

RA Duyk G., Schneider M.C., Geng L., Zhang F., Zhao Z., Torosian S.,
RA Reeders S.T., Bork P., Pohlschmidt M., Loehning C., Kraus B.,
RA Nowicka U., Leung A.L.S., Frischauf A.-M.;
RT "Polycystic kidney disease: the complete structure of the PKD1 gene
RT and its protein.";
RL Cell 81:289-298(1995).
RN [2]
RP SEQUENCE OF 2769-4303 FROM N.A.
RX MEDLINE=94273192; PubMed=8004675;
RA Ward C.J., Peral B., Hughes J., Thomas S., Gamble V.,
RA MacCarthy A.B., Sloane-Stanley J., Buckle V.J., Kearney L.,
RA Higgs D.R., Ratcliffe P.J., Harris P.C., Roeliffsema J.H.,
RA Spruit L.L., Saris J.J., Dauwerse H.G., Peters D.J.M.,
RA Breuning M.H., Nellist M., Brook-Carter P.T., Maheshwar M.M.,
RA Cordeiro I., Santos H., Cabral P., Sampson J.R., Janssen B.,
RA Hesselink-Janssen A.L.W., van den Ouweland A.M.W., Eussen B.,
RA Verhoef S., Lindhout D., Halley D.J.J.;
RT "The polycystic kidney disease 1 gene encodes a 14 kb transcript and
RT lies within a duplicated region on chromosome 16. The European
RT Polycystic Kidney Disease Consortium.";
RL Cell 77:881-894(1994).
RN [3]
RP STRUCTURE BY NMR OF 275-354.
RX MEDLINE=99107746; PubMed=9889186;
RA Bycroft M., Bateman A., Clarke J., Hamill S.J., Sandford R.,
RA Thomas R.L., Chothia C.;
RT "The structure of a PKD domain from polycystin-1: implications for
RT polycystic kidney disease.";
RL EMBO J. 18:297-305(1999).
RN [4]
RP VARIANT ADPKD 3748-ARG-VAL-3752 DEL, AND VARIANT ASP-3632.
RX MEDLINE=96108969; PubMed=8554072;
RA Peral B., San Millan J.L., Ong A.C.M., Gamble V., Ward C.J.,
RA Strong C., Harris P.C.;
RT "Screening the 3' region of the polycystic kidney disease 1 (PKD1)
RT gene reveals six novel mutations.";
RL Am. J. Hum. Genet. 58:86-96(1996).
RN [5]
RP VARIANTS ADPKD P-2993; R-3016 & V-3511, AND VARIANTS M-3510 & F-4190.
RX MEDLINE=97342914; PubMed=9199561;
RA Peral B., Gamble V., Strong C., Ong A.C.M., Sloane-Stanley J.,
RA Zerres K., Winearls C.G., Harris P.C.;
RT "Identification of mutations in the duplicated region of the
RT polycystic kidney disease 1 gene (PKD1) by a novel approach.";
RL Am. J. Hum. Genet. 60:1399-1410(1997).
RN [6]
RP VARIANT ALA-4058.
RX MEDLINE=97295081; PubMed=9150733;
RA Constantinides R., Xenophontos S.L., Neophytou P., Nomura S.,
RA Pierides A., Constantinou-Deltas C.D.;
RT "New amino acid polymorphism, Ala/Val4058, in exon 45 of the
RT polycystic kidney disease 1 gene: evolution of alleles.";
RL Hum. Genet. 99:644-647(1997).
RN [7]
RP VARIANTS T-2760;P-2761;V-2763;T-2764;Q-2791;T-2826;L-3008 & L-3064.
RX MEDLINE=97449169; PubMed=9285784;
RA Watnick T.J., Piontek K.B., Cordal T.M., Weber H., Gandolph M.A.,
RA Qian F., Lens X.M., Neumann H.P.H., Germino G.G.;
RT "An unusual pattern of mutation in the duplicated portion of PKD1 is
RT revealed by use of a novel strategy for mutation detection.";
RL Hum. Mol. Genet. 6:1473-1481(1997).
RN [8]
RP VARIANT ADPKD THR-3678.
RX MEDLINE=97403939; PubMed=9259200;
RA Turco A.E., Rossetti S., Bresin E., Englisch S., Corra S.,
RA Pignatti P.F.;
RT "Three novel mutations of the PKD1 gene in Italian families with
RT autosomal dominant polycystic kidney disease.";
RL Hum. Mutat. 10:164-167(1997).
RN [9]
RP VARIANT ADPKD ASP-4032, AND VARIANT VAL-4045.
RX MEDLINE=98180892; PubMed=9521593;
RA Daniells C., Maheshwar M.M., Lazarou L., Davies F., Coles G.,

```


RA Ravine D.;
RT "Novel and recurrent mutations in the PKD1 (polycystic kidney
RL disease) gene.";
RN Hum. Genet. 102:216-220(1998).
[10]
RP VARIANT ADPKD MET-3375.
RX MEDLINE-99118881; PubMed-9921908;
RA Koptides M., Constantinides R., Kyriakides G., Hadjigavriel M.,
RA Patsalis P.C., Pierides A., Deltas C.C.;
RT "Loss of heterozygosity in polycystic kidney disease with a missense
RT mutation in the repeated region of PKD1.";
RL Hum. Genet. 103:709-717(1998).
[11]
RN VARIANTS ADPKD L-324 AND VARIANTS R-1399 AND L-1786.
RX MEDLINE-99294580; PubMed-10364515;
RA Thomas R.L., McConnell R., Whittaker J., Kirkpatrick P., Bradley J.,
RA Sandford R.;
RT "Identification of mutations in the repeated part of the autosomal
RT dominant polycystic kidney disease type 1 gene, PKD1, by long-range
RT PCR.";
RL Am. J. Hum. Genet. 65:39-49(1999).
[12]
RN VARIANTS ADPKD L-3994--F-3996 DUPL; G-4136 AND C-4154, AND VARIANTS.
RX MEDLINE-20441957; PubMed-10987650;
RA Perrichot R.A., Mercier B., Simon P.M., Whebe B., Cledes J., Ferrec C.;
RT "DGE screening of PKD1 gene reveals novel mutations in a large cohort
RT of 146 unrelated patients.";
RL Hum. Genet. 105:231-239(1999).
[13]
RN VARIANTS ADPKD 3748-R-V-3752 DEL AND L-4132 DEL, AND VARIANT V-4045.
RX MEDLINE-20112427; PubMed-10847901;
RA Afzal A.R., Hand M., Ternes-Pereira E., Saggar-Malik A., Taylor R.,
RA Jeffery S.;
RT "Novel mutations in the 3' region of the polycystic kidney disease 1
RT (PKD1) gene.";
RL Hum. Genet. 105:648-653(1999).
[14]
RN VARIANTS ADPKD PRO-4225 AND TRP-4276.
RX MEDLINE-99217041; PubMed-10200984;
RA Badenas C., Torra R., San Millan J.L., Lucero L., Milla M.,
RA Estivill X., Darnell A.;
RT "Mutational analysis within the 3' region of the PKD1 gene.";
RL Kidney Int. 55:1225-1233(1999).
[15]
RN VARIANTS ADPKD MET-2250 AND TRP-2329, AND VARIANTS.
RX MEDLINE-20311156; PubMed-10854095;
RA Perrichot R., Mercier B., Quere I., Carre A., Simon P., Whebe B.,
RA Cledes J., Ferrec C.;
RT "Novel mutations in the duplicated region of PKD1 gene.";
RL Eur. J. Hum. Genet. 8:353-359(2000).
[16]
RN VARIANTS ADPKD PRO-2971 AND MET-3375, AND VARIANT LBU-3066.
RX MEDLINE-20382887; PubMed-10923040;
RA Koptides M., Mean R., Demetrou K., Constantinides R., Pierides A.,
RA Harris P.C., Deltas C.C.;
RT "Screening of the PKD1
RT nucleotide polymorphisms and a de novo mutation in Hellenic
RT polycystic kidney disease families.";
RL Hum. Mutat. 16:176-176(2000).
[17]
RN VARIANTS ADPKD GLN-3719 AND PRO-3852, AND VARIANT VAL-4045.
RX MEDLINE-20514565; PubMed-11058904;
RA Agiari G., Savelli S., Garbo M., Bozza A., Augello G., Penolazzi L.,
RA De Paoli Vitali E., La Torre C., Cappelli G., Piva R., del Senno L.;
RT "Novel splicing and missense mutations in autosomal dominant
RT polycystic kidney disease 1 (PKD1) gene: expression of mutated
RT genes.";
RL Hum. Mutat. 16:444-445(2000).
[18]
RN VARIANTS ADPKD, AND VARIANTS.
RX MEDLINE-20467506; PubMed-11012875;
RA Phakdeekitcharoen B., Watnick T.J., Ahn C., Whang D.-Y., Burkhardt B.,
RA Germino G.G.;
RT "Thirteen novel mutations of the replicated region of PKD1 in an Asian
RL population.";
RN Kidney Int. 58:1400-1412(2000).
[19]
RP VARIANTS ADPKD TRP-3753 AND ASN-3815.
RX MEDLINE-20275386; PubMed-10729710;
RA Kim U.K., Jin D.K., Ahn C., Shin J.H., Lee K.B., Kim S.H., Chae J.J.,
RA Wang D.Y., Lee J.G., Namkoong Y., Lee C.C.;
RT "Novel mutations of the PKD1 gene in Korean patients with autosomal
RT dominant polycystic kidney disease.";
RL Mutat. Res. 432:39-45(2000).
[20]
RN VARIANTS ADPKD, AND VARIANTS.
RX MEDLINE-21063179; PubMed-11115377;
RA Rossetti S., Strmecki L., Gamble V., Burton S., Sneddon V., Peral B.,
RA Roy S., Bakaloglu A., Komel R., Winearls C.G., Harris P.C.;
RT "Mutation analysis of the entire PKD1 gene: genetic and diagnostic
RT implications.";
RL Am. J. Hum. Genet. 68:46-63(2001).
CC -I- FUNCTION: PKD1 AND PKD2 MAY FUNCTION THROUGH A COMMON SIGNALING
CC PATHWAY THAT IS NECESSARY FOR NORMAL TUBULOGENESIS. INVOLVED IN
CC ADHESIVE PROTEIN-PROTEIN AND PROTEIN-CARBOHYDRATE INTERACTIONS.
CC -I- SUBUNIT: INTERACTS WITH PKD2.
CC -I- DISEASE: DEFECTS IN PKD1 ARE THE CAUSE OF AUTOSOMAL DOMINANT
CC POLYCYSTIC KIDNEY DISEASE (ADPKD), A COMMON AUTOSOMAL DOMINANT
CC GENETIC DISEASE AFFECTING ABOUT 1 OUT 1000 INDIVIDUALS. IT IS
CC CHARACTERIZED BY PROGRESSIVE FORMATION AND ENLARGEMENT OF CYSTS IN
CC BOTH KIDNEYS, TYPICALLY LEADING TO END-STAGE RENAL DISEASE IN
CC ADULT LIFE. CYSTS ALSO OCCURS IN THE LIVER AND OTHER ORGANS.
CC -I- SIMILARITY: CONTAINS 2 LEUCINE-RICH REPEATS (LRR).
CC -I- SIMILARITY: CONTAINS 16 PKD DOMAINS.
CC -I- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN (ATYPICAL,
CC THE POTENTIAL CALCIUM-BINDING SITE IS MISSING).
CC -I- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
Query Match 90.6%; Score 29; DB 1; Length 4303;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CSVTVG 6
Db 1270 CTVTVG 1275
RESULT 8
RK2_SOYBN
ID RK2_SOYBN STANDARD; PRT; 143 AA.
AC P18663;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CHLOROPLAST 50S RIBOSOMAL PROTEIN L2 (FRAGMENT).
GN RPL2.
OS Glycine max (Soybean).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-88143992; PubMed-3344206;
RA Spielmann A., Roux E., von Allmen J.M., Stutz E.;
RT "The soybean chloroplast genome: complete sequence of the rps19 gene,
RT including flanking parts containing exon 2 of rpl2 (upstream), but
RT rpl22 (downstream)."
RL Nucleic Acids Res. 16:1199-1199(1988).
CC -I- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X06429; CAA29735.1; -
 DR PIR; S00718; S00718.
 DR Mendel; 4043; GLYma:rp12;1.
 DR InterPro; IPR002171; Ribosomal_L2.
 DR Pfam; PF00181; Ribosomal_L2; 1.
 DR PROSITE; PS00467; RIBOSOMAL_L2; 1.
 KW Ribosomal protein; Chloroplast.
 FT NON_TER 1
 SQ SEQUENCE 143 AA; 15215 MW; 046F005D7E51D22A CRC64;

Query Match 87.5%; Score 28; DB 1; Length 143;
 Best Local Similarity 83.3%; Pred. No. 16;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CSVTVG 6
 |||||
 Db 55 CSATVG 60

RESULT 9
 SNAT_MACMU STANDARD; PRT; 204 AA.
 ID SNAT_MACMU
 AC 097756;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SEROTONIN N-ACETYLTRANSFERASE (EC 2.3.1.87) (ARYALKYLAMINE N-
 DE ACETYLTRANSFERASE) (AA-NAT) (SEROTONIN ACETYLASE).
 GN AANAT OR SNAT.
 OS Macaca mulatta (Rhesus macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OC NCBI_TaxID=9544;
 OX [1]

RN SEQUENCE FROM N.A.
 RP TISSUE=Pinel gland;
 RA Coon S.L.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CATALYZES THE N-ACETYLATION OF SEROTONIN INTO N-
 CC ACETYL-SEROTONIN.
 CC -1- CATALYTIC ACTIVITY: ACETYL-COA + ARALKYLAMINE = COA + N-
 CC ACETYLARALKYLAMINE.
 CC -1- PATHWAY: CONVERSION OF SEROTONIN TO MELATONIN; PENULTIMATE STEP.
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PINEAL GLAND AND RETINA.
 CC -1- SIMILARITY: BELONGS TO THE ACETYLTRANSFERASE FAMILY. AANAT
 CC SUBFAMILY.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U46661; RAD00061.1; -
 DR InterPro; IPR000182; Acetyltransf_GCN5.
 DR Pfam; PF00583; Acetyltransf; 1.
 KW Transferase; Acyltransferase.
 FT ACT_SITE 165
 FT ACT_SITE 165
 SQ SEQUENCE 204 AA; 22920 MW; 40BA615F6A9195E9 CRC64;

BY SIMILARITY.

Query Match 87.5%; Score 28; DB 1; Length 204;
 Best Local Similarity 66.7%; Pred. No. 22;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CSVTVG 6
 |::|||
 Db 176 CAITVG 181

RESULT 10
 SNAT_CHICK STANDARD; PRT; 205 AA.
 ID SNAT_CHICK
 AC P79774;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SEROTONIN N-ACETYLTRANSFERASE (EC 2.3.1.87) (ARYALKYLAMINE N-
 DE ACETYLTRANSFERASE) (AA-NAT) (SEROTONIN ACETYLASE).
 GN AANAT OR SNAT.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 OX [1]

RN SEQUENCE FROM N.A.
 RP TISSUE=Pinel gland;
 RX MEDLINE=9713331; PubMed=8978728;
 RA Bernard M.; Iuvone P.M.; Cassone V.M.; Roseboom P.H.; Coon S.L.;
 RA Klein D.C.;
 RT "Avian melatonin synthesis: photic and circadian regulation of
 RT serotonin N-acetyltransferase mRNA in the chicken pineal gland and
 RT retina."
 RL J. Neurochem. 68:213-224(1997).
 CC -1- FUNCTION: CATALYZES THE N-ACETYLATION OF SEROTONIN INTO N-
 CC ACETYL-SEROTONIN.
 CC -1- CATALYTIC ACTIVITY: ACETYL-COA + ARALKYLAMINE = COA + N-
 CC ACETYLARALKYLAMINE.
 CC -1- PATHWAY: CONVERSION OF SEROTONIN TO MELATONIN; PENULTIMATE STEP.
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ACETYLTRANSFERASE FAMILY. AANAT
 CC SUBFAMILY.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U46502; AAB40942.1; -
 DR InterPro; IPR000182; Acetyltransf_GCN5.
 DR Pfam; PF00583; Acetyltransf; 1.
 KW Transferase; Acyltransferase.
 FT ACT_SITE 166
 FT ACT_SITE 166
 SQ SEQUENCE 205 AA; 23187 MW; 40871147793A4D80 CRC64;

BY SIMILARITY.

Query Match 87.5%; Score 28; DB 1; Length 205;
 Best Local Similarity 83.3%; Pred. No. 22;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CSVTVG 6
 |::|||
 Db 177 CQTVG 182

RESULT 11
 SNAT_MOUSE STANDARD; PRT; 205 AA.
 ID SNAT_MOUSE
 AC O88816;
 DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE SEROTONIN N-ACETYLTRANSFERASE (EC 2.3.1.87) (ARYALKYLAMINE N-
DE ACETYLTRANSFERASE) (AA-NAT) (SEROTONIN ACETYLASE).
GN AANAT OR SNAT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Retina;
RX MEDLINE=98372593; PubMed=9708862;
RA Sakamoto K., Ishida N.;
RT "Molecular cloning of serotonin N-acetyltransferase gene from the
RT mouse and its daily expression in the retina.";
RL Neurosci. Lett. 250:181-184(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/HE, AND 129SV; TISSUE=Pineal gland;
RX MEDLINE=99058023; PubMed=9838107;
RA Roseboom P.H., Nambodiri M.A., Zimonjic D.B., Popescu N.C.,
RA Rodriguez I.R., Nambodiri M.A., Zimonjic D.B., Popescu N.C.,
RA Rodriguez I.R., Gastel J.A., Klein D.C.;
RT "Natural melatonin 'knockdown' in C57BL/6J mice: rare mechanism
RT truncates serotonin N-acetyltransferase.";
RL Brain Res. Mol. Brain Res. 63:189-197(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RA Roseboom P.H., Nambodiri M.A., Zimonjic D.B., Popescu N.C.,
RA Rodriguez I.R., Klein D.C.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE N-ACETYLATION OF SEROTONIN INTO N-
CC ACETYL-SEROTONIN.
CC -1- CATALYTIC ACTIVITY: ACETYL-COA + ARALKYLAMINE - COA + N-
CC ACETYLARALKYLAMINE.
CC -1- PATHWAY: CONVERSION OF SEROTONIN TO MELATONIN; PENULTIMATE STEP.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PINEAL GLAND AND RETINA.
CC -1- SIMILARITY: BELONGS TO THE ACETYLTRANSFERASE FAMILY. AANAT
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB013358; BAA31526.1; -;
DR EMBL; AF004108; AAD09408.1; -;
DR EMBL; U83462; AAD08637.1; -;
DR MGD; MGI:1328365; Aanat.
DR InterPro; IPR000182; Acetyltransf_GCN5.
DR Pfam; PF00583; Acetyltransf; 1.
KW Transferase; Acetyltransferase.
FT ACT_SITE 166 166
FT ACT_SITE 166 166 BY SIMILARITY.
SQ SEQUENCE 205 AA; 23068 MW; D2ECD070998CB643 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 205;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 177 CAITVG 182
|:|:|
|:|:|
RESULT 12
SNAT_HUMAN STANDARD; PRT; 207 AA.
ID SNAT_SNAT_HUMAN

AC Q16613;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SEROTONIN N-ACETYLTRANSFERASE (EC 2.3.1.87) (ARYALKYLAMINE N-
DE ACETYLTRANSFERASE) (AA-NAT) (SEROTONIN ACETYLASE).
GN AANAT OR SNAT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96299662; PubMed=8661026;
RA Coon S.L., Mazuruk K., Bernard M., Roseboom P., Klein D.C.,
RA Rodriguez I.R.;
RT "The human serotonin N-acetyltransferase (EC 2.3.1.87) gene (AANAT):
RT structure, chromosomal localization, and tissue expression.";
RL Genomics 34:76-84(1996).
CC -1- FUNCTION: CATALYZES THE N-ACETYLATION OF SEROTONIN INTO N-
CC ACETYL-SEROTONIN.
CC -1- CATALYTIC ACTIVITY: ACETYL-COA + ARALKYLAMINE - COA + N-
CC ACETYLARALKYLAMINE.
CC -1- PATHWAY: CONVERSION OF SEROTONIN TO MELATONIN; PENULTIMATE STEP.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PINEAL GLAND AND RETINA.
CC -1- SIMILARITY: BELONGS TO THE ACETYLTRANSFERASE FAMILY. AANAT
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U40391; AAC50555.1; -;
DR EMBL; U40347; AAC50554.1; -;
DR MIM; 600950; -;
DR InterPro; IPR000182; Acetyltransf_GCN5.
DR Pfam; PF00583; Acetyltransf; 1.
KW Transferase; Acetyltransferase.
FT ACT_SITE 168 168 BY SIMILARITY.
SQ SEQUENCE 207 AA; 23344 MW; 7476612F3661E0D5 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 207;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVTVG 6
Db 179 CAITVG 184
|:|:|
|:|:|
RESULT 13
YXHK_CAEEL STANDARD; PRT; 229 AA.
ID YXHK_CAEEL
AC Q20728;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 25.4 KDA PROTEIN F53F4.3 IN CHROMOSOME V.
GN F53F4.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Barlow K.;

```
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 UBIQUITIN-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE CKAP1/ALP1 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 277663; CAB01212.1; -.
DR WormPep; F53F4.3; CE10958.
DR InterPro; IPR000938; CAP-Gly.
DR Pfam; PF01302; CAP_GLY; 1.
DR PROSITE; PS00845; CAP_GLY_1; 1.
DR PROSITE; PS00503; UBIQUITIN_2; 1.
KW Hypothetical protein; Cytoskeleton; Microtubules.
FT DOMAIN ? UBIQUITIN-LIKE.
FT DOMAIN 170 212 CAP-GLY.
SQ SEQUENCE 229 AA; 25440 MW; C465365DAE378A0F CRC64;

Query Match 87.5%; Score 28; DB 1; Length 229;
Best Local Similarity 83.3%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVTWG 6
DB 152 CEVTWG 157

RESULT 14
RK2_NICDE STANDARD; PRT; 266 AA.
AC P21434;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CHLOROPLAST SUGAR-BINDING PROTEIN L2.
GN RPL2.
OS Nicotiana debneyi (Debnay's tobacco).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4089;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84297246; PubMed=6089120;
RA Zurawski G., Bottomley W., Whitfield P.R.;
RT "Junctions of the large single copy region and the inverted repeats
RT in Spinacia oleracea and Nicotiana debneyi chloroplast DNA: sequence
RT of the genes for trnA and the ribosomal proteins S19 and L2."
RL Nucleic Acids Res. 12:6547-6558(1984).
CC -!- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X00798; CAB52367.1; -.
DR PIR; S07356; R5NT2D.
DR Mendel; 5309; NICDE; rpl2.1.
DR InterPro; IPR002171; Ribosomal_L2.
DR Pfam; PF00181; Ribosomal_L2; 1.
DR PROSITE; PS00467; RIBOSOMAL_L2; 1.

KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 266 AA; 28901 MW; 8F4AAE3CE2ADF314 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 266;
Best Local Similarity 83.3%; Pred. No. 29;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVTWG 6
DB 186 CSATVG 191

RESULT 15
NIVA_CLOPA STANDARD; PRT; 269 AA.
ID NIVA_CLOPA
AC Q00853;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE HOMOCITRATE SYNTHASE, ALPHA SUBUNIT (EC 4.1.3.21).
GN NIFV-ALPHA.
OS Clostridium pasteurianum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1501;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91216974; PubMed=2022611;
RA Wang S.-Z., Dean D.R., Chen J.-S., Johnson J.L.;
RT "The N-terminal and C-terminal portions of Nifv are encoded by two
RT different genes in Clostridium pasteurianum."
RL J. Bacteriol. 173:3041-3046(1991).
CC -!- FUNCTION: THIS PROTEIN IS A FE-MO-COPFACTOR BIOSYNTHETIC
CC COMPONENT.
CC -!- CATALYTIC ACTIVITY: 2-HYDROXYBUTANE-1,2,4-TRICARBOXYLATE + COA =
CC ACETYL-COA + H(2)O + 2-OXOGLUTARATE.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND AN OMEGA CHAIN.
CC -!- MISCELLANEOUS: IN CLOSTRIDIUM PASTEURIANUM THE N-TERMINAL AND
CC C-TERMINAL PORTIONS OF NIFV ARE ENCODED BY TWO DIFFERENT GENES,
CC NIFV-ALPHA AND NIFV-OMEGA.
CC -!- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE
CC SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M68878; AAA23268.1; -.
DR PIR; B39403; B39403.
DR InterPro; IPR002034; AIPM_homocit_synth.
DR Pfam; PF00682; HMGL-like; 1.
DR PROSITE; PS00815; AIPM_HOMOCIT_SYNTH_1; 1.
DR PROSITE; PS00816; AIPM_HOMOCIT_SYNTH_2; 1.
KW Nitrogen fixation; Lyase.
SQ SEQUENCE 269 AA; 29864 MW; 782C60A6595A3979 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 269;
Best Local Similarity 83.3%; Pred. No. 29;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVTWG 6
DB 128 CEVTWG 133
```

Search completed: January 31, 2002, 09:36:53
Job time: 772 sec

THIS PAGE BLANK (uspto)